



# Thinking in Patterns

Fractals and Related Phenomena in Nature

Editor

Miroslav M. Novak

# **Thinking in Patterns**

Fractals and Related Phenomena in Nature



This page intentionally left blank

# Thinking in Patterns

Fractals and Related Phenomena in Nature



Editor

**Miroslav M. Novak**

School of Mathematics, Kingston University, UK

 **World Scientific**

NEW JERSEY • LONDON • SINGAPORE • SHANGHAI • HONG KONG • TAIPEI • CHENNAI



*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**THINKING IN PATTERNS**

Copyright © 2004 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-238-822-2

Printed in Singapore by Mainland Press

to *Erika, Karl and Doreen*



This page intentionally left blank

## Contents

<b>Preface</b>	<i>M. M. Novak</i>	<b>xi</b>
<b>Selected Topics in Mathematics, Physics, and Finance Originating in Fractal Geometry</b>	<i>B. B. Mandelbrot</i>	<b>1</b>
<b>A Renewal Process of Mittag-Leffler Type</b>	<i>F. Mainardi, R. Gorenflo and E. Scalas</i>	<b>35</b>
<b>On the Activity of Absorbing Irregular Interfaces</b>	<i>J. S. Andrade Jr., H. F. Da Silva, E. A. Henrique and B. Sapoval</i>	<b>47</b>
<b>Fractal Deformation Using Displacement Vectors and Their Increasing Rates Based on Extended Unit Iterated Shuffle Transformation</b>	<i>T. Fujimoto and N. Chiba</i>	<b>57</b>
<b>Multifractal and Stochastic Analysis of Electropolished Surfaces</b>	<i>M. Haase, A. Mora and B. Lehle</i>	<b>69</b>
<b>A Method for Numerical Estimation of Generalized Rényi Dimensions of Affine Recurrent IFS Invariant Measures</b>	<i>T. Martyn</i>	<b>79</b>
<b>Nonlinear Dynamics and Prediction of the Caspian Sea Level</b>	<i>N. G. Makarenko, L. M. Karimova, Y. B. Kuandykov and M. M. Novak</i>	<b>91</b>
<b>Self-similarity in Plants: Integrating Mathematical and Biological Perspectives</b>	<i>P. Prusinkiewicz</i>	<b>103</b>
<b>Cognitive Scale-free Networks as a Model for Intermittency in Human Natural Language</b>	<i>P. Allegrini, P. Grigolini and L. Palatella</i>	<b>119</b>
<b>The Complexity of Biological Ageing</b>	<i>D. Stauffer</i>	<b>131</b>
<b>Fitting Curves by Fractal Interpolation: An Application to the Quantification of Cognitive Brain Processes</b>	<i>M. A. Navascues and M. V. Sebastian</i>	<b>143</b>



<b>Stochastic and Regular Components in Forcing of Solar Large-scale Structures</b> <i>E. Tikhomolov</i>	155
<b>Fast, Efficient On-line Simulation of Self-similar Processes</b> <i>O. D. Jones</i>	165
<b>Fractal Geometry in the Arts: An Overview Across The Different Cultures</b> <i>N. Sala</i>	177
<b>Fractal Properties and Characterization of Road Profiles</b> <i>P. Legrand, J. Lévy Véhel and M.-T. Do</i>	189
<b>Fractal Distributions of Temperature, Salinity and Fluorescence in Spring 2001-2002 in South San Francisco Bay</b> <i>K. Fisher and W. Kimmerer</i>	199
<b>Characterization of Fractal Structures Through a Hausdorff Measure Based Method</b> <i>F. Nekka and J. Li</i>	213
<b>Fractal Scattering Indicators for Urban Sound Diffusion</b> <i>P. W. Woloszyn</i>	221
<b>Binomial Multiplicative Model of Critical Fragmentation</b> <i>H. Katsuragi, D. Sugino and H. Honjo</i>	233
<b>Study on the Improved Fractal Interpolation Surface of the Attitude and Surface of Fault</b> <i>H. Sun and H. Xie</i>	243
<b>A Deterministic Power Domain Algorithm for Fractal Image Decompression</b> <i>N. Nikolaou, A. Kakos and V. Drakopoulos</i>	255
<b>Comparative Dynamical Scaling Analysis of Quasi-2D Electrodeposited Silver Patterns under Localized and Non-localized Random Quenched Noise</b> <i>M. A. Pasquale, S. L. Marchiano and A. J. Arvia</i>	267
<b>Epidermal Ridges: Positional Information Coded in an Orientational Field</b> <i>M. B. Nguyen, V. Fleury and J.-F. Gouyet</i>	279
<b>Multiscale Principal Components</b> <i>A. Saucier</i>	291

<b>Coexistence of Doublon and Dendrite Structure with Phase-Field Model</b> <i>S. Tokunaga and H. Sakaguchi</i>	<b>301</b>
<b>Fractality and Fractal Dimension in Mesoamerican Pyramid Analysis</b> <i>G. Burkle-Elizondo, A. G. Fuentes-Larios and R. D. Valdez-Cepeda</i>	<b>309</b>
<b>Morphological Variety in Crystal Growth of Mercury (II) Chloride on Agar Slides</b> <i>J. A. Betancourt-Mar and E. J. Suárez-Domínguez</i>	<b>311</b>
<b>Fractal Characteristics of Bainbridge Crater Lake Sediment Gray-scale Intensity Data Documenting the Frequency and Intensity of Holocene El Niño/Southern Oscillation Events</b> <i>N. A. Bryksina and W. M. Last</i>	<b>313</b>
<b>Fractals and Plant Water Use Efficiency</b> <i>A. Bari, G. Ayad, A. Martin, J. L. Gonzalez-Andujar, M. Nachit, and I. Elouafi</i>	<b>315</b>
<b>Need and Feasibility of Applying L-system Models in Agricultural Crop Modeling</b> <i>L. Pachepsky, M. Kaul, Ch. Walthall, C. Daughtry and J. Lydon</i>	<b>317</b>
<b>Fractal Detection and Avoidance Using RS Statistics and Honeybee Navigational Skills in Dynamic Environments</b> <i>R. L. Walker</i>	<b>319</b>
<b>Signal and Image Processing with FracLab</b> <i>J. Lévy Véhel and P. Legrand</i>	<b>321</b>
<b>Author Index</b>	<b>323</b>

This page intentionally left blank

## Preface

The term fractal was conceived by Benoit Mandelbrot some 30 years ago. Since then, the new area of fractal geometry has been developed and it has quickly penetrated the established areas of practically all the disciplines of the natural and life sciences. This non-Euclidean geometry united the often disparate scientific fields by demonstrating a common underlying thread that throws new light on the classification and characterization of events and processes. This modern approach has gained universal approval in recent years.

This book, *Thinking in Patterns*, celebrates Benoit B. Mandelbrot's 80<sup>th</sup> birthday and his momentous contribution to the field of fractal geometry. Every reader interested in the broad field of nonlinear phenomena will find something of interest here. The abundance of papers and the range of topics appearing here confirm the underlying similarity between subjects such as

finance	cognitive processes	fluctuations of sea levels
road profiles	biological aging	solar magnetic fields
sound diffusion	electropolished surfaces	arts across cultures
image decompression	epidermal ridges	

The papers in this book are based on presentations at the 8<sup>th</sup> international conference, *Fractal 2004*, exploring the above mentioned issues. The conferences are now regular and well established among the nonlinear series of conferences and provide a unique and genial atmosphere to foster exchange and incubation of ideas. This travelling conference series is organized in different geographical regions to encourage international collaboration. Among the many distinguishing features of this series is its multidisciplinary nature, which has been growing steadily.

*Fractal 2004* was made possible through the help of the following members of the programme committee (in alphabetical order):

P. Allegrini (*Italy*), J. S. Andrade Jr (*Brazil*), M. Daoud (*France*), K. Falconer (*UK*), M. Frame (*USA*), J.-F. Gouyet (*France*), R. Hilfer (*Germany*), A. Holden (*UK*), S. Lovejoy (*Canada*), F. Mainardi (*Italy*), E. Mosekilde (*Denmark*), U. Parlitz (*Germany*), M. M. Novak (*UK - Chair*), W. H. Steeb (*South Africa*)

The *Fractal 2004* conference was partially supported by Emergentis Ltd.

The details about the next conference in this series will be posted on the web site <http://www.kingston.ac.uk/fractal/>.

M. M. Novak  
Kingston-upon-Thames, UK

This page intentionally left blank

# SELECTED TOPICS IN MATHEMATICS, PHYSICS, AND FINANCE ORIGINATING IN FRACTAL GEOMETRY

BENOIT B. MANDELBROT

*Mathematics Department, Yale University, New Haven, CT 06520-8283, USA*  
*email: benoit.mandelbrot@yale.edu*

The bulk of this text consists in nonsystematic sketches of the current status of diverse very difficult questions in various mathematical sciences. All were triggered by actual fractal pictures generated by computer. In physics some of those questions outline a nascent “rational rugometry,” involving quantitative measures of roughness. Other questions concern diverse clusters and turbulence. In mathematics, some of those questions have been settled — one of them, the  $4/3$  conjecture, in 2000. Other questions, however, including a basic property of the Mandelbrot set, resist repeated efforts to answer them. In finance, Mandelbrot’s models starting in 1963 became the foundation of “econophysics.” In all cases, many questions on the research frontier — solved or not — can be understood by a good secondary-school student, which is why fractal geometry is increasingly affecting high school teaching. All those questions involve in essential fashion some shapes long called “monsters” and guaranteed to belong to esoteric mathematics lacking any contact with the real world. Fractal geometry reveals them as being extremely “natural” and also as having been familiar to artists since time immemorial.

## 1 Introductory comments of various kinds

### 1.1 Presentation

Fractal geometry ranges over many parts of the mathematical sciences but the questions sketched in this text mostly belong to either pure mathematics or the interfaces between mathematics and physics. Specific sections or subsections are free-standing and do not require acquaintance with one another or with fractal geometry as a whole.

The paper may also interest those already familiar with fractal geometry because it includes recent developments and/or because many of my opinions have either evolved or become more focussed. Hence — even though the overall tone is by no means introductory — it is appropriate to begin with several separate introductory remarks concerned, first, with science, then with mathematics.

### 1.2 Dilation invariance and a reinterpretation of fractal geometry, as the first step towards a “quantitative rational rugometry”

A basic issue must be touched first: what is fractal geometry today? Largely after the fact, it is best characterized as being the first systematic and quantitative approach to the study of roughness — in both in pure mathematics and in mathematical sciences of the “real world.” The latter includes nature (turbulence, clusters of statistical physics, broken solids, noises, galaxy distributions, geomorphology) and “culture,” that is, the works of Man (finance, spoken discourse, the internet, and even art).

Roughness is, of course, ubiquitous in the real world and has long been counted among the basic “sensation” of Man. However, its study lagged; even finding a

quantitative measure for it was a challenge that defied an easy answer. Science was powerless to tackle roughness until I found that in many cases it obeys diverse geometric scaling rules that can be accounted for by a dilation invariance. Fractures of metals are iconic from that viewpoint, as pointed out in Section 1.4.

Three forms of dilation invariance stand out. A fractal whose detailed structure is a reduced-scale image of the overall shape (perhaps statistically deformed), is called “self-similar.” When the reduction ratios are different in different directions, the fractal is “self-affine.” When the reduction ratios vary from point to point, one deals with “multifractality” (Section 3.3)

A first key continuing part of fractal geometry consists in identifying and classifying cases ruled by some form of dilation invariance.

A second key continuing part of fractal geometry results from the fact that dilation invariance provides the study of roughness with an increasing number of intrinsic quantitative tools — beginning with several distinct flavors of fractal dimension. That is, dilation invariance is the ingredient that makes roughness manageable. This is also why fractal geometry is a very broadly useful first approximation. Rough aspects of mathematics, nature, and culture come together because they can be studied by closely related tools, and progress in each aspect benefits from progress in the others. But unity stops at a certain point: each example has specific features that must eventually be acknowledged.

In 1975, having conceived and began to develop systematically a nascent geometry of roughness, I turned to the Latin adjective for “rough and broken up,” namely *factus*, and coined for this geometry the term *fractal*.

Let me now restate the key scientific claim I put forward increasingly forcibly and continue in buttressing. A workable path towards rational rugometry has now been identified as being made of rough shapes that are dilation invariant. They are the fractals.

### 1.3 *Explanatory background in older sciences that study other sensations of Man*

It is good to keep in mind that the earliest sciences started as ways to organize substantial collections of messages that Man receives from the various “senses.” The complexity of most messages is such that a science can take off only after it identifies “representative” special cases to be studied first.

For acoustics, an important step consisted in recognizing that chirps or drums are very difficult to handle, but idealized vibrating strings or pipes lead to periodic sums of sinusoids. That is, acoustics became quantitative when it managed to define “pure sounds” and measure their pitch by a frequency. As had to be the case, this quantitative measure is consistent with “intuition” and the extensive earlier knowledge manifested, for example, in music. The limitations of acoustics continue to be notorious, but do not prevent it from being extraordinarily useful.

Similarly, the theory of heat became quantitative when Galileo devised the thermometer and measured hotness by a temperature. Here too, a limitation must be recognized: far from equilibrium, the theory of heat continues to struggle.

In the same vein, the examples of real rough curves or surfaces that are usefully

close to being self-similar or self-affine allowed me to define “pure” or “perfect” roughness as analogous to the classic concept of “perfect gas in equilibrium.” The latter is invariant by translation of time, the fractals — once again — by dilation.

Like pure sound or pure elliptic motion under gravitation, pure roughness is an abstraction and fractal geometry cannot address roughness that is far from being dilation-invariant. But dilation-invariant roughness is useful: its scope is considerable and must be expanded before facing further tasks.

#### 1.4 *Fractal dimension as the first intrinsic and quantitative measure of roughness; metal fractures and a conjectured fundamental universality*

As first measures of pure roughness, I proposed notions that were known but viewed as esoteric: fractal dimension or Hölder exponent or codimension. It was necessary to first reinterpret these notorious concepts as being numerical characteristics of an invariance (self-similarity, self-affinity, or multifractality) and then expand their study, both concretely and intuitively.

From the preceding viewpoint, particular iconic importance attaches to a study by myself, Passoja & Paullay (*Nature*, 308 (1984) 721-2). We found metal fractures to be dilation invariant with a dimension that exceeds 2 — the dimension of smooth surfaces — by  $1/3$ . This property has been confirmed by extensive later work that went beyond metals to glasses and covered sizes covering five decades at least. The range is sometimes even broader, but may be limited by the nature of the data. Fractality is the special ingredient making it possible to measure roughness intrinsically by what is now often called the “roughness exponent.”

This discovery of the “universal” excess dimension  $1/3$  has provided the nascent rational rugometry of metal fractures with a broad and fundamental observation. It defines a “macroscopic” aspect of the study of fracture that must be added as conjecture to the more prevalent “microscopic” approaches.

An invidious claim one hears is that fractal geometry has solved or advanced no existing problem in physics. This claim is, among others, contradicted in the contexts of metal fractures and turbulence. But it may be true that the more visible role of fractals in physics has not been directed to what already existed but to the future. The very fact of proposing a quantitative measure of roughness has raised *thoroughly new problems* of all kinds. Several have already been solved, for example problems concerning the fractal dimensions of two very distinct kinds of physical clusters, examined in Sections 2.6 and 6.3. Other new problems remain wide open and there is no reason to expect them to be easy.

#### 1.5 *A fundamental formal kinship between the nascent “rational rugometry” and thermometry*

The suitable measure of roughness having been found in previously esoteric notions of mathematics, rugometry might have developed in ways quite distinct from the sciences based on previously quantified “sensations.” But in important cases fractal dimension takes the form  $\Sigma p \log p$ , which is an “information” hence a further link with thermodynamical entropy. This resemblance is far from complete but brings a high level of formal unity and suffices to allow many questions concerning roughness to be handled by a near-thermodynamical formalism.



### 1.6 *Regrettable “centrifugal” tendencies splitting the fractal synthesis. The many historically separate notions of “scaling”*

Today — to my great regret — “centrifugal” developments affect several “chapters” of my work that arose in the 1950s and 1960s. All had been slow in acquiring a broad following until they were empowered by being subsumed in fractal geometry. Now they have taken off and tend to develop on their own. Some are commented upon in suitable sections of this paper. One is the study of Zipf’s and other “power laws” and Lévy stable distributions, which I began in the 1951. Another is “econophysics,” which I originated in 1962 without giving it any specific name. A third is the study of metallic fractures and the like. If these developments “dismember” the fractal synthesis, the resulting fragments would all be harmed.

Neither is it helpful to replace the term “fractal” by “scaling.” That replacement is sometimes formally correct but is invariably misleading because scaling has multiple meanings — related but not identical. Scaling occurs in probability theory since Cauchy (1853) and P. Lévy (the 1920s). It occurs in turbulence since Richardson (the 1920s) and Kolmogorov (1941). It occurs in increasingly geometric fashion in my work, since 1951 for Zipf’s law, and already very explicitly in 1956.

Finally, scaling occurs in different parts of “core physics,” especially in the physics of criticality since K. Wilson in 1972. Criticality had the largest number of practitioners and tempts other investigations to use its terminology. However, criticality is a very specific situation. The study of critical shapes like clusters have been greatly helped by fractal tools but there was no significant influence in the opposite direction. Not only criticality played no role in originating the chapters of fractal geometry mentioned early in this subsection, but it evolved no tools to help their study. For example, it had no use for Lévy stable distributions. Therefore, thinking in terms of criticality did not and does not bring any benefit.

Added to other reasons, the preceding comments make it useful to ponder the broader issue of the place of fractals within physics. I think of fractality as related to the emergence of a new stream of thinking sketched in Section 1.1. Being concerned specifically with roughness in all its forms, it can be viewed as providing a generalized physics. The dream of generalizing physics in this fashion is an ancient one but had long been thwarted as long as overly specific features of existing physics were preserved too faithfully.

### 1.7 *The role of fractal geometry in pure mathematics: renewed key role played by the “material” world and the examination of fully-fledged pictures*

Another invidious claim one hears is that fractal geometry has solved no existing mathematical problem. This claim has no merit, either, but it is true that I provided few difficult proofs but many separate conjectures of all kinds. Each turned out to be difficult and opened a new field that continues and prospers long after I move to other concerns. Notable examples will be mentioned in Section 7 devoted to the Mandelbrot set, Section 2 devoted to the dimension  $4/3$ , and several subsections throughout devoted to multifractality. Other conjectures are scattered elsewhere in this text.

The perceived importance of those contributions to pure mathematics varies

but a common feature is that they did not arise from earlier mathematics but in the course of practical investigations into diverse sciences of nature or of culture, some of them old and well-established, others newly revived, and a few altogether new. Some branches of mathematics agree that physics, numerical experimentation and geometric intuition are very beneficial but other branches proclaim physics as irrelevant, computation as powerless, and intuition as misleading — especially when it is strongly visual. A well-known irony is that history consistently proves that, as branches of mathematics develop, they suddenly either lose or acquire deep but unforeseen connections with the sciences — old and new.

As to numerical experimentation — which Gauss had found invaluable, but whose practice was long interrupted — it has seen its power multiplied thanks to computer calculations, and later, to computer graphics. This allows my practice to be dominated, in mathematics as in the sciences, by the role played by fully-fledged pictures that are as detailed as possible and go well beyond mere sketches and diagrams.

This feature destroyed a belief that was near-universal among pure mathematicians around 1980, that a picture can only lead to another, and never to fresh mathematics. Hence, my work bears on an issue of great consequence. Does pure (or purified) mathematics exist as an autonomous discipline, one that can — and ideally should — adhere to a Platonic ideal and develop in total isolation from both “sensations” and the “real” world? I believe, to the contrary, that the existence of totally pure mathematics is a myth — a useful one on occasion, but not on the long run.

My 1982 book *The Fractal Geometry of Nature*, *FGN*, was meant above all to be a “manifesto” in praise of the trained eye. I believe that computer graphics has changed the iconoclastic (anti-pictorial) dogma that prevailed in mathematics and physics into a serious liability. In search of always fresh evidence for this belief, I looked for new facts that the standard pictures leave hidden. The pictures’ original goal was modest: to gain acceptance for ideas and theories that I had managed to develop without pictures and whose acceptance was reluctant and slow because of cultural gaps. To begin with, the pictures did indeed lead to acceptance, but then they went on to help me and many others generate new ideas and theories. The input of mundane questions gradually grew and became far more ambitious than originally intended or recognized.

Norbert Wiener once described his key contribution to science as bringing together — starting from widely opposite horizons — the fine mathematical points of Lebesgue integration and the physics of Gibbs and Perrin. Similarly unlike “parents” characterize the theory of fractals, which is arguably a multiple second flowering of Wiener’s Brownian motion. Also (like Poincaré) Wiener was very committed (and successful) in making frontier science known to a wide public.

### 1.8 *The unexpectedly long history of fractals began well before nineteenth century mathematics; fractals have now been traced back to art since time immemorial*

Anticipating the difficult conjectures mentioned in Section 1.7, the early pictures I drew of old standbys like the Koch or Peano curves and the Cantor set were precise, and as a result they became inspiring. They sufficed to thoroughly disprove the previously held belief that those sets are “monsters.” Quite to the contrary, they were turned around into unavoidable “cartoons” of reality. For example, I “demoted” Peano “curves” from being counter-intuitive monsters to being nothing but motions that follow plane-filling networks of rivers.

More profoundly, giving concrete uses to mathematics allowed it to be compared on more equal terms with other human activities and allowed fractals’ history to slowly reveal itself as having been long and varied.

In art and decoration, they have been known since time immemorial, all over the world. I noted a few examples in *FGN* but new examples reveal themselves continually.

Far better known is the already mentioned second broad stage in history: a century ago, fractals entered the purest of mathematical esoterica and a “Polish school” of mathematics viewed itself as devoted exclusively to *Fundamenta*, added mightily to the list of monster shapes. It greatly contributed to the deep and long — but inevitably of finite duration — estrangement of mathematics from physics.

Specifically ironical, therefore, is that in a third stage my work, that of my colleagues, and now that of many scholars, made those monster shapes, and new shapes that are even more “pathological,” into everyday tools of the sciences of nature and culture.

This subsection must end by a call for balance. I always agreed with John von Neumann that “a large part of mathematics which became useful developed with absolutely no desire to be useful... This is true for all science. Successes were largely due to... relying solely on... intellectual elegance. It was by following this rule that one actually got ahead in the long run, much better than any strictly utilitarian course would have permitted... The principle of *laissez-faire* has led to strange and wonderful results.”

### 1.9 *The beauty of fractals*

Fractal pictures have become ubiquitous. Many strike everyone as being of exceptional and totally unexpected beauty. Some have the beauty of the mountains and clouds they are meant to represent; others are abstract and seem wild and unexpected at first, but after brief inspection appear totally familiar. In front of our eyes, the visual geometric intuition built on the practice of Euclid and of calculus is being retrained with the help of new technology.

Hence a different philosophical issue arises. Is there any relation between the beauty of these mathematical pictures and the beauty that a mathematician rooted in the twentieth century mainstream sees in his trade after long and strenuous practice? My lectures often underline these questions, by showing in full colors what certain mathematical shapes really look like.

### 1.10 General references

Due to space restrictions, this survey is extremely sketchy and centers around my own contributions. As the field grew, early versions appeared in 1984, 1999, 2000, and 2001. Each in turn was made obsolete by the continuing development of the field.

On fractals overall, the basic reference remains my 1982 book *The Fractal Geometry of Nature*, already referenced as *FGN*. As explained at the end of the paper, suitable other initials in italics will reference other books, some printed and others only available (now or shortly) on my web: [www.math.yale.edu/mandelbrot](http://www.math.yale.edu/mandelbrot). More specific references are made part of the text.

Alternative surveys include a) a text I wrote with M.L. Frame for *The Encyclopedia of Physical Science and Technology in Fifteen Volumes* (San Diego CA: Academic), third edition (2001): 6, 185-208, b) the Overview chapter of *SH*, and c) several chapters of book *MF*. A useful commentary on the mathematics is provided by the Foreword Peter W. Jones contributed to *SC*.

## 2 Complex Brownian bridge; Brownian cluster and the dimension $4/3$ of its boundary; the self-avoiding plane Brownian motion

The sequence of examples in this paper follows little order. As mathematics goes, the iconic Mandelbrot set is only mentioned in Section 7. The present Section 2 is concerned with an example that is far less widely known but is easy to understand and of greatest current interest. It provided mathematicians with difficult conjectures and a unifying theme. It provided physics with a new cluster having special virtues discussed in Section 2.5.

### 2.1 A historically incorrect and continuing misleading “streamlined” story

The story of the “ $4/3$  conjecture” was exemplary by the standards of my work and this paper but very atypical by the customary standards of mathematics. Therefore it is often replaced by the following grossly “streamlined” account.

Somehow, Mandelbrot had the idea that in the plane the boundary of Brownian motion is a curve of Hausdorff-Besicovitch dimension  $4/3$ . The conjecture attracted wide attention but turned out to be very challenging. The proof took time and came in two stages.

A “field-theoretical” physical argument has been provided by B. Duplantier, *Phys. Rev. Lett.* 82, 1999, 880; 82, 1999, 3940; 84, 2000, 1363.

A proper proof has been provided by G. Lawler, O. Schramm & W. Werner, much of it is only available on the Web ([xxx.lanl.gov/abs/math.PR/0010165](http://xxx.lanl.gov/abs/math.PR/0010165)) as a series of preprints totaling over 100 pages, the first of which has been accepted by *Acta Mathematica*. According to a newsweekly (*Science*, 8 December 2000, pages 1883-4) it “drew rave reviews” at an important meeting and was hailed as “one of the finest achievements in probability theory in the last 20 years.”

Between 1982 and 2000, a dozen or so scattered technical conjectures in mathematical analysis had been shown to be equivalent to that “ $4/3$ .” Therefore, all have now been proven as corollaries and together provide an element of unity that continues to be explored.

## 2.2 Preliminaries to the historical sequence of events. Definitions of the Brownian cluster and of self-avoiding Brownian motion

The actual history of the 4/3 is more interesting. The key discovery reported in 1982 — *FGN*, Plate 243 — relied on a novel processing of *Brownian motion*  $B(t)$  in the plane. This very old shape is, of course, a random process whose increments  $B(t+h) - B(t)$  are two-dimensional Gaussian random variables with mean 0 and variance  $h$ , and are independent over disjoint time intervals. It is well-known that  $B(t)$  is statistically self-affine in the sense that

$$\Pr\{B(t+h) - B(t) \leq b\} = \Pr\{B(s(t+h)) - B(st) \leq \sqrt{s}b\},$$

and the same is true of joint probability distributions for all finite collections of time intervals  $h_j$ .

Assuming  $B(0) = 0$ , a *Brownian bridge*  $B_{\text{bridge}}(t)$  was defined by N. Wiener as the periodic function of  $t$ , of period  $2\pi$ , that is defined for  $0 \leq t \leq 2\pi$  by

$$B_{\text{bridge}}(t) = B(t) - (t/2\pi)B(2\pi).$$

In distribution,  $B_{\text{bridge}}(t)$  is identical to a sample of  $B(t)$  conditioned to return to  $B(0) = 0$  for  $t = 2\pi$ . Wiener wrote  $B_{\text{bridge}}(t)$  as a trigonometric series whose  $n$ th coefficient is  $G_n/\sqrt{n}$ , where the  $G_n$  are independent reduced Gaussian random variables. Combining two statistically independent Brownian bridges yields the complex function  $B_{\text{bridge}}(t) = B_r(t) + iB_i(t)$ .

The *Brownian plane cluster*  $Q$  is defined in *FGN*, Plate 243, as the set of values of  $B_{\text{bridge}}(t)$ . This is the (non-traditional) map of the time axis by the complex function  $B_{\text{bridge}}(t)$ . The classical map of the time axis by  $B(t)$  is everywhere dense in the plane, and the map of a time interval by  $B(t)$  is an inhomogeneous set. In contrast, conditioning the origin  $\Omega$  of the frame of reference to belong to  $Q$  makes all the probability distributions concerning  $Q$  independent of  $\Omega$ . Therefore  $Q$  (see *SN*, Chapters 8, 9 and 10) I called  $Q$  a *conditionally homogeneous* set. This property is not only aesthetically attractive, but, as will be seen, proved inspiring.

The *self-avoiding planar Brownian motion*  $\tilde{Q}$ . This random object is defined in *FGN* as being the closed set of points  $P$  in  $Q$  accessible from infinity by a path that does not intersect  $Q - P$ . This  $\tilde{Q}$  is also conditionally homogeneous.

## 2.3 Steps that led to the Brownian cluster being defined

Today, after the fact, the boundary of Brownian motion or cluster seems a “natural” notion. After all, the overall appearance of planar Brownian motion is known at least since J. Perrin, as evidenced in *FGN*, Plate 13. It inspired Norbert Wiener in the 1920s, then pictures’ evocative power was exhausted. In the absence of suitable “graphic rendering,” the earlier pictures of samples of  $B(t)$  did not highlight a boundary. Worse, they gave no hint of anything worth studying.

This boundary came up during a “fishing expedition,” an aimless search motivated by the feeling that a careful fresh look at  $B(t)$  using better tools may lead to new insight. Plate 242 of *FGN* exemplifies the finite duration samples of  $B(t)$  with which I began; those pictures “did not talk to me.” I figured that those finite samples’ non-homogeneity may overwhelm and hide interesting facts. When the eye

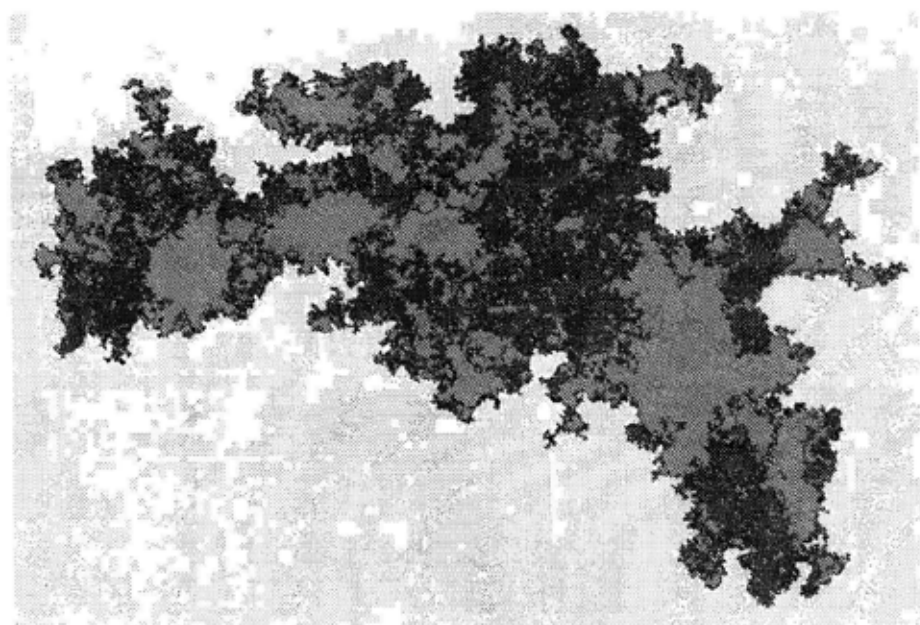


Figure 1. This is Plate 243 of *FGN*, representing the original sample of Brownian cluster.

is to be trusted, it is good practice to help it and in particular to avoid burdening it by extraneous complications — such as non-homogeneity.

To the contrary, the Brownian cluster is homogeneous by design. Therefore, I asked my assistant to produce a Brownian cluster and also to “paint” its interior in order to enhance the graphics.

The outcome became Plate 243 of *FGN*, reproduced here as Figure 1. It triggered an “eurêka” moment. With no prompting, what I saw looked to me like an island with a clearly visible and especially wiggly coastline. Hence visual intuition nourished by experience in geomorphology suggested  $D \approx 4/3$ . This value was confirmed by my direct numerical tests.

#### 2.4 *Comment on the relation between the dimension $4/3$ and self-avoidance*

Originally, the term “self-avoiding Brownian motion” came to my mind because  $\tilde{Q}$  is a shape related to Brownian motion and does not self-intersect. The term became strengthened because I recalled the dimension  $4/3$  found in the plane for the self-avoiding random walk (SARW) on a lattice. The value  $4/3$  for SARW is unquestioned but physicists obtained it by analytic arguments that are geometrically opaque; its interpretation as a dimension implies yet another unproven conjecture, which no one doubts.

## 2.5 Differences between the self-avoiding Brownian motions defined by the cluster and via the “streamlined” account

The mathematicians who take the “shortcut” described in Section 2.1 define “self-avoiding Brownian motion” as the boundary of a finite sample of Brownian motion. The same Hausdorff Besicovich dimension of  $4/3$  holds for two clearly distinct fractal curves. I suspect that the cluster boundary is the more interesting topic.

This ambiguity recalls one that specialists in SARW on a lattice have observed long ago: a standard definition and a “true” one had vied for attention. This topic is interesting but space lacks to develop it.

## 2.6 Brownian clusters, as compared to the clusters of statistical physics

Section 5 will survey several major clusters in statistical physics: percolation, Ising, DLA. All belong to physics on a prescribed lattice. Contrary to fractals, their construction does not proceed by an interpolation that converges strongly to a limit, but by extrapolation.

It is the case that down-scaled versions of those physical lattice clusters, converge weakly to fractals? This is what I conjectured and precise forms of the conjectures are widely believed and studied. For DLA (Section 5.3) the issue is murky.

By contrast, Brownian clusters did not originate in physics but have a special asset they follow an explicit definition and involve no conjectural limit process.

## 2.7 Squigs and a wide open issue that combines fractals and topology

Being obtained by extrapolation, SARW is difficult to study. In the spirit of Section 2.6, *FGN* (Chapter 24) introduced recursive alternatives to SARW, called *squigs*, that create self-avoidance by interpolation. For the simplest squig my heuristic argument yielded the dimension  $\log_2 2.5 \approx 1.3219\dots$  This value was confirmed by J. Peyrière, *C.R. Acad. Sc. Paris*: 286A 1978, 937 and *Ann. Institut Fourier*: 31, 1981, 187. The discrepancy between  $4/3$  and  $\log_2 2.5$  clearly follows from the fact that only the squigs — not the clusters — involve a discrete and recursive subdivision of the plane into triangles, squares, or other indefinitely interpolable tessellations. Viewing this discrepancy as of secondary importance, I suspect that self-avoidance is linked in a profound and intrinsic way to the dimension  $4/3$ . The nature of this link is a mystery and a challenge.

## 3 Explosive multiplication of new fractal constructions, dimensions (including negative ones), and Hölder exponents

Until fractal geometry became organized, the numbers of distinct fractal constructions and of distinct definition of fractal dimension were both very small. Moreover, the values of distinct dimensions used to coincide, except for contrived “counterexamples.” As fractals became common tools in the sciences and favorites in computer graphics new constructions multiplied. Moreover, differences between the values of distinct dimensions ceased to be exceptional; in many contexts they became the rule with every variant contributing its share to an overall description. Fractional



Brownian motion and multifractal measures led to a rich mathematical literature that is exemplified in *SH* and *SN*, respectively. Other new constructions are less well known. Section 3.1 describes one example. The remainder of the chapter tackles the multiplicity of dimensions.

### 3.1 *A promising but little-explored novelty: embedding the stable processes and fractional Brownian motions in a the broader class of functions: the fractal sums of pulses*

Brownian motion was generalized in two deeply different ways by the introductions of Lévy stable processes (LSM) and fractional Brownian motions (FBM). The LSM depend on a parameter  $\alpha$ , with  $0 < \alpha \leq 2$  and  $\alpha = 2$  yielding the Brownian as a limiting case. They are investigated, among many other places, in *SE*. The FBM depend on a parameter  $H$ , with  $0 < H < 1$  and  $H = 1/2$  yielding the Brownian as a critical case. They are investigated, among many other places, in *SH*. By the definition of  $B_H(t)$ , the increment  $B_H(t) - B_H(t')$  is a Gaussian random variable of expectation 0 and standard deviation  $|t - t'|^H$ .

Numerous formal analogies exist between the respective studies of LSM and FBM. Those analogies changed from surprising to very natural when I imbedded both families in a far broader family, the “fractal sums of pulses” (FSP). The FSP also allow a variety of additional behaviors that are useful in science and may be of mathematical interest. The latest reference is my contribution to *Long-Range Dependent Processes* (eds. G. Rangarajan and Ming Ding) Springer 2003, pp. 118-135.

### 3.2 *Multiplicity of alternative definitions of dimension*

*Linearly self-similar sets are iconic but exceptional. For them, the many definitions of fractal dimension yield identical values.* A set  $S$  is self-similar if it is constructed recursively and its generator consists of  $N$  copies of itself, the  $i$ th copy  $S_i$  being obtained from  $S$  by a similarity with contraction factor  $r_i$ . The calculation of the fractal dimension  $D$  is relatively simple. Under a mild condition (the “open set” condition),  $D$  is the solution of the Moran generating equation

$$\sum r_i^D = 1,$$

where  $i$  ranges from 1 to  $N$ .

The original Hausdorff-Besicovitch dimension invoked in Section 2 remains essential in mathematics despite the fact that its value is often hard to obtain. But in the sciences,  $D_{BH}$  is impossible to measure because its definition contains the operation “inf.” (In the case of self-similar or self-affine shapes, the operation “limit” poses no problem.) Far more important is the fact that self-similar sets are a special case. Purely mathematical needs demanded concepts of dimension distinct from  $D_{HB}$  and contrived “counter-examples” showed that, in the absence of self-similarity, those dimensions can take distinct values. More recently, concrete needs forced fractal geometry to alternative definitions that led to values other than  $D_{HB}$ . Often, considering those values together helps describe an object’s geometry.



### 3.3 Self-similar multifractal measures

The random multiplicative singular measures that I began to construct around 1970 are described in papers from 1968, 1972, 1974 and 1976 collected in *SN*. They are now called multifractal. They were not intended to become a new kind of esoterica but a model in turbulence and (near immediately after) in finance. The conjectures I put forward created an active and prosperous subbranch of mathematics and — today — the main branch of statistical modelling of the variation of financial prices.

The topic is too rich to be dwelt upon here, but it is useful to note that a multifractal measure is, above all, described by a function  $f(\alpha)$  of the parameter  $\alpha$ . My original 1974 paper dealt with multiplicative multifractals (see Section 3.5) and deduced a function equivalent to  $f(\alpha)$  from the Cramer theory of large deviations. Since they involve a function  $f(\alpha)$ , multifractal measures involve an infinite number of parameters.

### 3.4 Negative dimensions as measure of the newly introduced notion of quantitative measure of emptiness

The value of  $f(\alpha)$  can be either  $\geq 0$  or  $\leq 0$ , hence a fundamental distinction enters inevitably. When it is positive,  $f$  is a suitable set's fractal dimension, for example in the sense of Hausdorff Besicovitch. When it is negative,  $f$  takes an altogether different new role, as a measure of "degree of emptiness." (Mandelbrot, *J. Fourier Analysis and Applications* (Kahane issue), 1995, 409-432; *J. Stat. Physics*, 110, 2003, 739-777). Negative dimensions amply deserve closer study.

### 3.5 Multiplicative multifractals: microcanonical, canonical, and products of pulses or other functions

Multifractals' self-affinity can be approximate or exact. Numerous approaches, some heuristic and some mathematically rigorous, apply under quite general conditions but, as unavoidable counterpart, they are not very specific. Beginning in my pioneer papers, I have taken a different tack and deliberately focussed on multifractals that — in a statistical sense — are exactly self-affine. They are less general but perspicuous and continue to yield very specific and varied results one can "tune" by changing the process.

Step by step, the constraints were made less and less strong and immensely richer structures arose. In 1974, I moved the construction from microcanonical to canonical products (*J. Fluid Mechanics* 62, 1974, 331-358 and CR (Paris) 278A; 1974, 289-292 & 355-358). Recently, the construction moved further to products of pulses and of other kinds of functions (Barral and Mandelbrot *Proba. Th. and Related Fields* 124, 2002, 409-430, *J. Math. Pures et Appl.* 82, 2003, 1555-1589 and contributions to the book *Fractals* (ed. M. Lapidus) Am. Math. Soc., 2004.)

### 3.6 Self-affine sets

When the transformation of  $S$  into  $S_i$  is an affinity, the evaluation of  $D_{HB}$  was successful in a surprisingly small number of cases. Contributors include McMullen,

Bedford, Falconer, Peres, Kenyon, Lalley, and Gatzouras.

Furthermore, the many alternative definitions of fractal dimension yield values that differ from  $D_{HB}$  and from one another. In particular, my contribution to *Fractals in Physics* (E. Pietronero & E. Tosatti, eds.) 1986 (reprinted in *SH* as Chapters H22, H23 and H24) introduced the concepts of *local* and *global* dimension. They coincide in the self-similar case but greatly differ in the case of self-affinity. The global notions of dimension pose many open mathematical issues.

All these computations suggest that, while the notion of fractal dimension can be defined under wide conditions, its “natural domain” of practical relevance centers around self similarity.

### 3.7 The many forms of the Hölder (and Hurst) exponent

In the case of the graph of a self-affine function, the most “natural” quantitative description of roughness is not provided by a dimension, but by diverse forms of an exponent introduced in the 1970s by the mathematicians by Hölder and Lipschitz and in the 1950s by the hydrologist H. E. Hurst. The variable  $\alpha$  in the multifractal function  $f(\alpha)$  is a Hölder exponent. Chapter E6 of *SE* and Chapter N1 of *SN* show that the original definitions have, in response to concrete needs, branched in diverse directions.

### 3.8 The exponent yielded by a generalized Moran equation

As discussed in *SE* and mentioned in Section 10.3, I put forward the fractional Brownian motions of multifractal “trading time” as models of price variation. Instead of a Hölder-Hurst exponent, they involve “ $H$ ” exponents of particularly great variety.

Denoting the  $\Delta P_i$  the increments of such a function over arbitrarily chosen time increments  $\Delta t_i$ , the sum  $|\Delta P_i|$  has no upper bound, hence  $P(t)$  is called a function of unbounded variation. More generally, define the  $q$ th variation by starting from the formula for the ordinary variation and replacing  $|dP|$  by  $|dP|^q$ . If the  $q$ th variation is infinite for  $q < 1/H$  and vanishes for  $q > 1/H$ , the value  $q = 1/H$  is “critical” and defines the tau dimension  $D_\tau$ . (The tau dimension is independent of the trading time and concern a projection along the time axis of a complex-valued “completion” of the function  $P(t)$ .) The inverse  $1/D_\tau$  is yet another form of Hölder’s exponent. It generalizes to all processes and in many cases the equation yielding  $D_\tau$  is a generalization of Moran’s equation of Section 3.2.

This is, for example, the case for the “cartoons” that I described in *Quantitative Finance*, 1, 2001, 427-440.

The properties of  $D_\tau$  and of the “non-Hölderian”  $1/D_\tau$  deserve careful mathematical study beyond what is already known.

## 4 Tools of fractal analysis other than the dimensions: ramification and lacunarity

Careful analysis brings in many fractal tools, some new, other old but obscure, that are neither dimension-like nor Hölder like exponents.

#### 4.1 Sierpinski curves and Urysohn-Menger ramification

As seen in *FGN*, Sierpinski's investigations in the 1900s built on two ancient decorative designs: one became known as the "carpet," and the second I called the "gasket." The Sierpinski carpet shows that a plane curve can be "topologically universal," that is, contain a (homeomorphic) transform of every other plane curve. The construction starts with a square, divides it into nine equal subsquares and erases the middle one, which I call a "trema" ( $\tau\rho\eta\mu\alpha$  is the Greek term for "hole"). One proceeds in the same fashion with each remaining subsquare, and so on ad infinitum. The Sierpinski gasket is a curve with branching points everywhere. The construction starts with an equilateral triangle, divides it into four equal subtriangles and erases the middle one as trema. One proceeds in the same fashion with each remaining subtriangle, and so on ad infinitum.

During the 1920s, the distinction between the carpet and the gasket became essential to the theory of curves. Piotr Urysohn and Karl Menger took them as prime examples of curves having, respectively, an infinite and a finite "order of ramification."

*FGN* quotes influential mathematicians for whom the "gasket" gave prime evidence that geometric intuition is powerless, because it can only conceive of branch points as being isolated, not everywhere dense. In fact, Gustave Eiffel himself wrote (as I interpret him) that he would have made his Tower lighter, with no loss of strength, had the cost of finer materials allowed him to increase the density of double points. From the Eiffel Tower to the Sierpinski gasket is an intellectual step that one's intuition is easily trained to take.

The theory of curves that studies carpets, gaskets and the order of ramification became a stagnant corner of mathematics. Where can one find the latest facts about these notions? The surprising answer is that, after I introduced them in the statistical physics of condensed matter, physicists came to view these notions as "unavoidable." Once ridden of the cobwebs of abstraction, they prove to be very practical and enlightening geometric tools to work with. Physicists make them the object of scores of articles, and invent scores of generalizations that mathematicians did not need in 1915.

#### 4.2 Ramification's key role in diffusion on fractals

Early on in the study of fractals in physics (in the wake of Gefen et al *Phys. Rev. Lett.*: 45, 1980, 855) we had to investigate random walks on lattices that approximate fractals. We found that a key role is played by those fractals' order of ramification. The theory was easy for  $R < \infty$  (for example for the Sierpinski gasket). But for  $R = \infty$  (for example, for the Sierpinski carpet), exact theory is impossibly difficult and we had to resort to possibly dubious approximations.

The theory of diffusion on fractals has grown into an active field of mathematics. For  $R < \infty$ , our heuristic arguments have been given a sound basis but the case  $R = \infty$  continues to be very problematic.

### 4.3 A non-dimensional and non-topological fractal tool that begs to be studied further: lacunarity

The well-known standard construction of a Cantor dust proceeds recursively as follows. The “initiator” is the interval  $0, 1$ . Its first stage ends with a generator made of  $N$  subintervals, each of length  $r$ . In the second stage, each generator interval is replaced by  $N^2$  intervals of length  $r^2$ , etc. The resulting limit set arose in the study of trigonometric series, but first attracted wider interest because of its topological and measure-theoretical properties. From those viewpoints, all Cantor dusts are equivalent. Hausdorff’s and every other definition of dimension yield  $D = \log N / \log(1/r)$ . The value of dimension splits the topological Cantor dusts into finer classes of equivalence parameterized by  $D$ .

Fractal geometry showed those classes of equivalence to be of great concrete significance. In due time, the needs of science, rather than mathematics, required an even finer subdivision. To pose a problem, consider the Cantor-like constructions stacked in Figure 2. In the middle line,  $N = 2$  and  $r = 4^{-1}$ ;  $k$  steps below the middle line,  $N = 2^k$ ,  $r = 4^{-k}$  and the generator intervals are uniformly spaced;  $k$  steps above the middle line,  $N = 2^k$ ,  $r = 4^{-k}$ , again, but the generator intervals are crowded close to the endpoints of  $0, 1$ . The Cantor dusts in this stack share the common values  $D = 1/2$ , but look totally different. The Latin word for hole being *lacuna*, motion down the stack (or up) is said to correspond to decreasing (increasing) *lacunarity*.

*Challenge.* As  $k \rightarrow \infty$ , the bottom line becomes “increasingly dense” in  $0, 1$ , and the top line “increasingly close to two dots.” Provide a mathematical characterization of this “singular” passage to the limit.

*Second challenge.* FGN, Chapters 33 to 35, and my contribution to *Fractal Geometry and Stochastics* (ed. C. Bandt et al) Birkhäuser 199, 12-38 describe and illustrate several constructions that allow a control of lacunarity. However, for the needs of both mathematics and science, the differences between the resulting constructs must be quantified. The existing studies of this quantification show that it is not easy and also not unique. Identical reduction ratios, like in Figure 2, create special complications.

Of the alternative methods investigated in the literature, one is based on the prefactor relation  $M(R) = FR^D$  that yields the mass  $M(R)$  contained in a ball of radius  $R$ .

Another method is based on the prefactor in the Minkowski content.

A third method has the advantage that defines a neutral level of lacunarity that separates positive and negative levels.

On the line, this level is achieved by any randomized Cantor dust  $S$  with the following property. Granted that any choice of origin  $\Omega$  in  $S$  divides the line into a right and a left half line, lacunarity is said to be neutral when the intersections of  $S$  by those half lines are statistically independent. Increasingly positive (resp. negative) correlations are used to express and measure increasingly low (resp. high) levels of lacunarity. These notions will be used in the sections that follow and in Section 6.3.

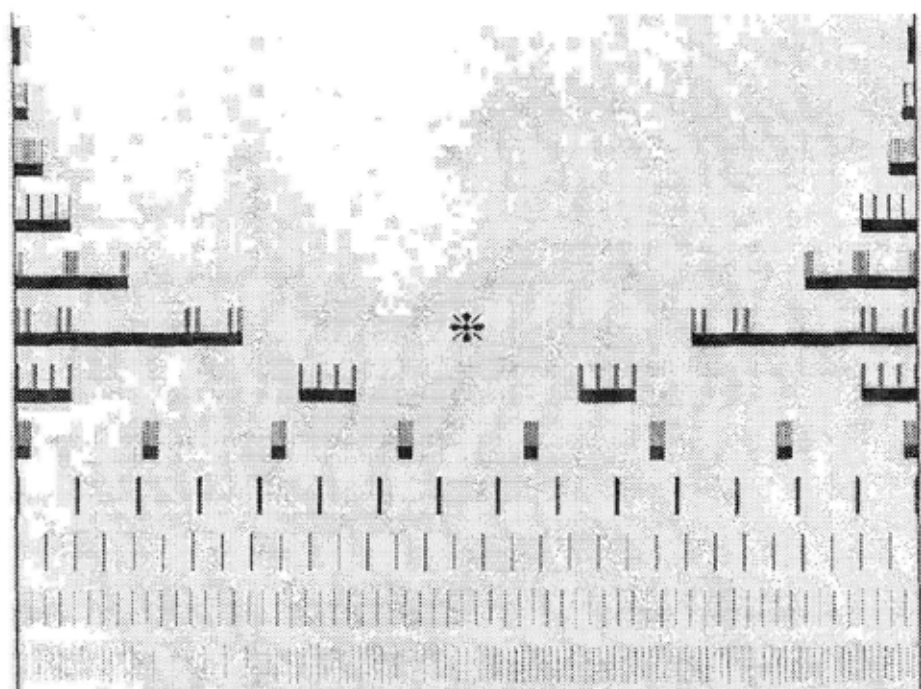


Figure 2. A stack of Cantor sets of equal dimension  $D = 1/2$ , whose lacunarity changes from very low at the bottom to very high at the top of the stack.

#### 4.4 *Actual geometric implementation of the formal fractional-dimensional spaces that are useful in statistical physics*

The physics of criticality is very successful with spaces whose properties are obtained from those of Euclidean spaces by interpolation to "noninteger Euclidean dimensions." The dimension may be  $4 - \epsilon$  or  $1 + \epsilon$ , where  $\epsilon$  is in principle infinitesimal. Formal calculations are carried out, including expansions in  $\epsilon$ . Then the final stage sets the "infinitesimal"  $\epsilon$  to  $\epsilon = 1$ . Mathematically, these spaces remain unspecified, yet the procedure turns out to be extremely useful.

*Mathematical challenge:* Show that the properties postulated for those spaces are mutually compatible, show that they do (or do not) have a unique implementation; describe their implementation constructively.

*Very partial solution:* A very special example of such space has been implemented as a limit (FGN, second printing, p. 462; Gefen et al, *Phys. Rev. Lett.* 50, 1983, 145). We showed that the postulated properties of certain physical problems in this space are identical to the *limits* of the properties of corresponding problems in a Sierpinski carpet whose "lacunarity" is made to converge to 0, in the sense that it tends to 0 as one moves down the stack on Figure 2.

## 5 Fractality of the major fractal clusters in statistical physics

While Brownian motion is fundamental in physics as well as in mathematics, the Brownian clusters of section 2 are recent, perhaps only a mathematical curiosity. However, their property of fractality is shared by all the major real clusters (turbulence, galaxies, percolation, Ising, Potts) and all the major real interfaces (turbulent jets and wakes; metal and glass fractures discussed in Section 1.4; diffusion fronts). Each of these categories raises numerous open mathematical questions, of which a few will be commented upon.

### 5.1 Percolation clusters at criticality

Take an extremely large lattice of copper or vinyl tiles. Each tile is chosen at random: with the probability  $p$ , it is made of vinyl and with the probability  $1 - p$ , of copper. Allow electric current to flow between two tiles if they have a side in common. A "cluster" can then be defined as a collection of copper tiles such that electricity can flow between any two of these tiles. The basic reference is D. Stauffer & A. Aharony. *Introduction to Percolation Theory*. Second edition. London: Taylor & Francis, 1992.

For an alternative, but equivalent, construction, define at the center of every tile a random "relief function"  $R(p)$  whose values are independent random variables uniformly distributed from 0 to 1. If this relief is flooded up to level  $p$ , each cluster stands out as a connected "island." Physicists conjectured, and mathematicians eventually proved, that there exists a "critical probability" denoted by  $p_C$ , such that a connected infinite island, or connected infinite conducting cluster, almost surely exists for  $p < p_C$ , but not for  $p > p_C$ .

The geometric complexity of percolation clusters at criticality is extreme, and many of the basic new conjectures did not arise from pure thought, but from careful examination of computer-generated clusters of unusually large size.

*Open conjecture A.* Take an increasingly large lattice and resize it to be a square of unit side. At  $p_C$ , the infinite cluster converges weakly to a "limit cluster" that is a fractal curve.

*Conjecture B.* The fractal dimension of this limit cluster is  $91/48$ . This value was first obtained numerically, then confirmed by den Nijs, from a partly heuristic "field theoretical" argument that yields characteristic exponents, finally made rigorous by S. Smirnov.

*Conjecture C.* Figure 3 shows that, depending on the definition of the boundary of a percolation cluster, its fractal dimension is either  $4/3$  or  $7/4$ . These conjectures began with experiments (Grossman and Voss, respectively) and rigorous mathematical proof have been provided by S. Smirnov.

It may be worth mentioning that proofs concerning fractal dimensions have attracted wider interest among mathematicians than the rigorous proofs of previously known facts about percolation.

*Open conjecture D.* Linear cross-sections of the limit cluster are Lévy dusts, as defined in *FGN*. Experimental evidence is found in Mandelbrot & Stauffer, *J. Physics: A* 28, 1995, L 213 and Hovi et. al. *Phys. Rev. Lett.*: 77, 1996, 877.

*Open conjecture E.* The limit cluster is a finitely ramified curve in the sense of

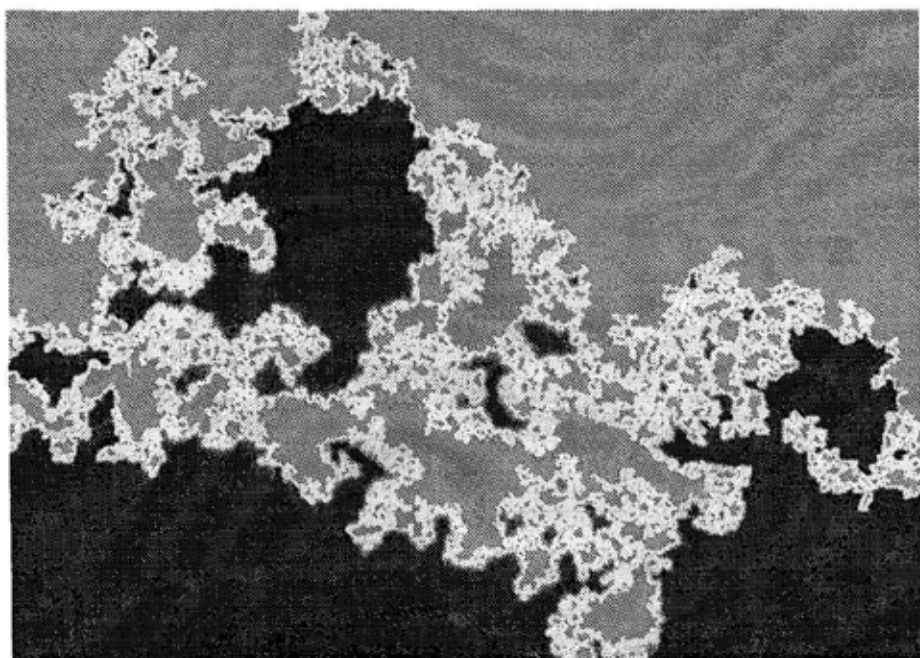


Figure 3. This figure (drawn by Bernard Sapoval for a different goal) helps explain why the critical percolation clusters have two sharply distinct boundaries. One is the curve drawn in white. It is the common boundary of the black and grey areas that it separates. It is very convoluted but without self-contact and its fractal dimension is  $7/4$ . But there are many points where it nearly self-contacts so that it creates "pores," and plugging the pores one defines a "boundary of boundaries" of dimension drastically reduced to  $4/3$ .

Urysohn-Menger.

## 5.2 The Ising model of magnets at the critical temperature

At each node of a regular lattice, the Ising model places a spin that can face up or down. The spins interact via forces between neighbors left to themselves, these forces create an equilibrium (minimum potential) situation in which all the spins are either up or down. However, a second input is added: the system is in contact with a heat reservoir, and heat tends to invert the spins. When the temperature  $T$  exceeds a critical value  $T_C$ , heat overwhelms the interaction between neighbors. For  $T < T_C$ , local interactions between neighbors overwhelm heat and create global structures of greatest interest.

My work touched upon several issues in the shape of the up (or down) clusters at criticality.

*Long open implicit question:* Beginning with Onsager, it is known that in Euclidean space  $R^E$  the necessary and sufficient condition for magnets to exist is that  $E > 1$ . There are innumerable mathematical differences between the  $R_E$  for  $E = 1$  and  $E > 1$ . Identify differences that matter for the existence of magnets.



*Partial answer.* The specific examples of the Sierpinski curves and of related fractal lattices suggest that magnets can exist when and only when the order of ramification is infinite. *FGN*, p. 139; Gefen et al, *Phys. Rev. Lett.*: 45, 1980, 855).

*Conjecture.* The above answer is of general validity.

*Unanswered challenge.* Rephrase the criterion of existence of magnets from the present and highly computational form, to a direct form that would give a chance of proving or disproving the preceding conjecture.

### 5.3 The ever-mysterious clusters of diffusion-limited aggregation (DLA)

A DLA cluster is generated by allowing an “atom” to perform Brownian motion starting far away until it hits an initial “seed.” In Figure 4, the seed is the (opened up) bottom of a half cylinder. When the atom and the seed hit, they are “fused,” and a fresh Brownian atom is launched against the enlarged target.

Overwhelming evidence from computer simulations shows that the arrival of many atoms transforms the seed into a cluster that shows about the same high degree of complexity at all scales of observation. Hence any mathematical definition of the concept of fractal must be constrained to include DLA.

The simplicity of the growth rules the DLA and its basic role in understanding many physical phenomena have motivated extensive quantitative studies. However, a full theory even a more informal understanding of the resulting complex structure are lacking. Over many orders of magnitude, the circle of radius  $R$  centered on the original and contains a mass  $M(R) \approx R^D$  with  $D = 1.715$ . But there are definite divergences from strict self-similarity — as seen for example in my paper in *Physica A* 191, 1992, 95-107 and my paper with Kol and Aharony in *Phys. Rev. Lett.* 88, 2002, 055501-1-4.

At an early stage, those deviations were thought to be no worse than those relative to critical phenomena. The latter has a well-developed theory, and it was hoped that a theory of DLA could be achieved in the absence of a careful and complete description. This optimistic view is no longer widely held, and a careful description cannot be neglected.

## 6 Interrelations between fractality and smooth variability: some cases may have a common origin in the usual partial differential equations

### 6.1 An apparent quandary: are smoothness and fractality doomed to coexist with no interaction?

To establish the presence of fractals in nature and culture was a daunting task to which a large portion of *FGN* is devoted. New and often important examples keep being discovered, but the hardest present challenge is to discover the *cause*, or more probably, the *causes* of fractality.

Some cases are reasonably clear. Thus, in the case of the percolation and Ising clusters in Section 5, fractality is the geometric counterpart of scaling and renormalization, that is, of the fact that the analytic properties of those objects follow a wealth of analytic “power-law relations.” Many mathematical issues, some of



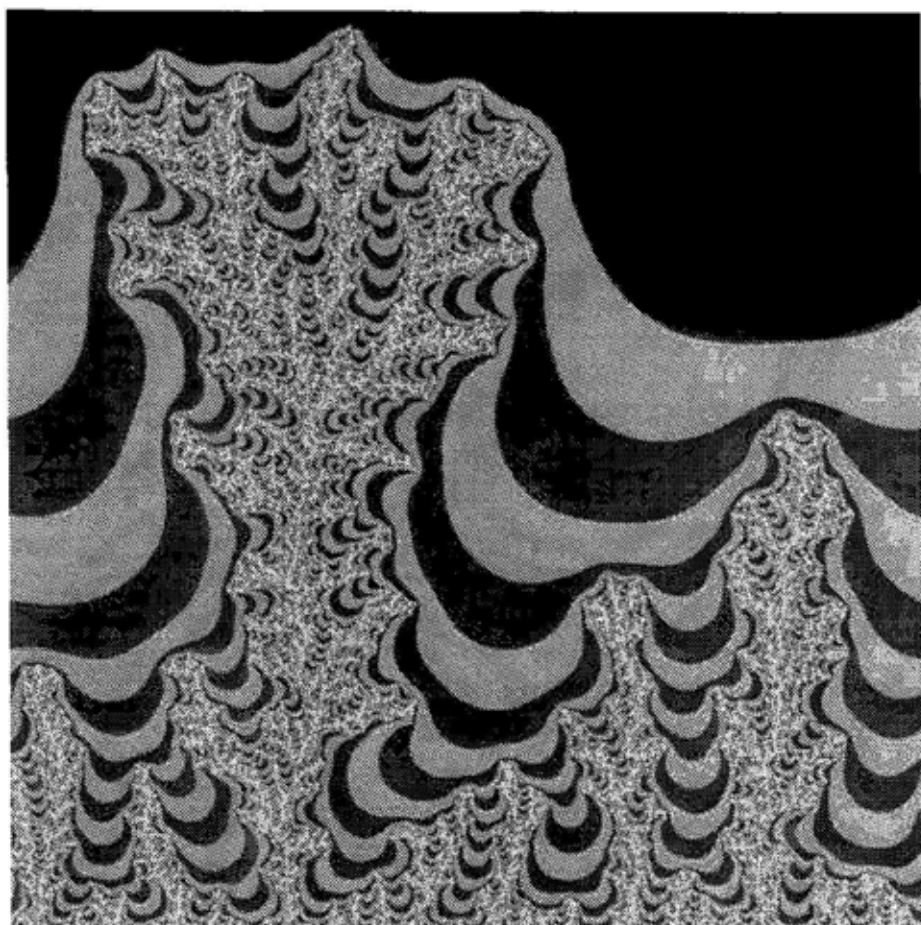


Figure 4. Reproduction of Figure C19-2 of *SC*. A smallish sample of plane DLA, called "cylindrical" because it is grown from the bottom of a half-cylinder (opened up). This DLA is small enough to compute the Laplacian potential and draw its isolines. The latter are a graphic device but also much more: an essential tool of study. A curious visual resemblance is thereby created between DLA and the Mandelbrot set. Of the two, DLA has proven the more resistant to analysis.

them already mentioned, remain open, but the overall renormalization framework is firmly rooted.

Renormalization and the resulting fractality also occur in arguments that involve the attractors and repellers of dynamical systems in a phase space. Best understood is renormalization for quadratic maps. Feigenbaum and others considered the real case. For the complex case, renormalization establishes that the Mandelbrot set (see Section 7) contains infinitely many small copies of itself.

Unfortunately, the usual renormalization fails — even in principle — to account for the diffusion-limited aggregates (DLA) and additional examples of fractality.

Yet another class important occurrences of fractality, to which we now proceed

is linked to partial differential equations in real space. It is universally granted that physics is ruled by diverse partial differential equations, PDEs. Those of Laplace, Poisson, and Navier-Stokes will be referred to as “basic.” All differential equations imply a great degree of local smoothness, even though closer examination shows isolated singularities or “catastrophes.” To the contrary, fractality implies everywhere dense (or at least widespread) roughness and/or fragmentation. This is one of the several reasons why fractal models in diverse fields were initially perceived as being “anomalies” that stand in direct contradiction with one of the firmest foundations of science.

### 6.2 *A conjecture stated and defended in FGN: the solutions of PDEs can be fractal*

This is no longer a conjecture, insofar as many specialized PDEs have been solved and found to create fractality. To eliminate the appearance of contradiction between smoothness and fractality, *FGN* conjectured that the same is true of the “basic” equation. This implies that fractals arise unavoidably in the long time behavior of the solution of very familiar and “innocuous”-looking equations. In particular, many concrete situations where fractals are observed involve equations having free and moving boundaries, and/or interfaces, and/or singularities.

As a suggestive “principle,” *FGN* (Chapter 11) described the possibility that, under broad conditions that largely remain to be specified, these free boundaries, interfaces and singularities converge to suitable fractals. Among equations examined from this viewpoint, this paper will limit itself to two examples of critical importance. In the case of DLA (Section 5.3), this argument supports self similarity, hence is disappointing, thus far.

### 6.3 *The large scale distribution of galaxies: Newton’s law as a possible sufficient generator of fractality*

*Background.* The near universally held view is that the distribution of galaxies is homogeneous, except for local deviations.

In the past, however *FGN*, Chapter 9, Y. Baryshev & P. Teerikorpi, *Discovery of Cosmic Fractals* Singapore: World Scientific 2002) a number of philosophers or science fiction writers have played with the notion that stars (galaxies were not known) follow a hierarchical distribution patterned — long in advance! — along a spatial Cantor set. Those models are excessively regular and necessarily imply that the Universe has a center assuming hierarchies leads to no prediction, that is, implies no property that was not put in beforehand, and raises no new question. For them and other good reasons, hierarchies were dismissed as unrealistic and largely forgotten.

*Conjecture that the distribution of galaxies is properly fractal.* *FGN*, Chaps. 9, 33, 34, and 35.) Granted that the distribution of galaxies certainly deviates from homogeneity, existing improvements took two broad approaches.

One consists in correcting for local inhomogeneity by using local “patches.”

My next simplest approach acknowledges that one must exclude strict hierarchies as being both physically unrealistic and in conflict with widely held principles.

But I also contend that the specific details of the hierarchical arguments are unimportant. What matters is the underlying fractality, which must be recognized as being of central importance and broad scope. To dismiss fractality with the hierarchies amounted to throwing the baby with the water.

To buttress this belief, I performed detailed mathematical and visual investigations of sample sites generated by two concrete constructions of random fractal sets. The details are given in *FGN*.

The first construction is *The Seeded Universe* that I based on a Lévy flight. Its Hausdorff-dimensional properties were well-known. I observed that its correlation properties (Mandelbrot, *C. R. Acad. Sc. Paris*: 280 A, 1975, 1075) are nearly identical to those of actual galaxy maps. The second construction is *The Parted Universe*, which is obtained by subtracting from space a random collection of overlapping sets, tremas.

In a statistical model, the self-similarity ratio is not restricted to powers of a prescribed  $r_0$ . That is, a hierarchical structure is not a deliberate and largely arbitrary input. Quite to the contrary, either of the above constructions yields sets that are highly irregular and involve no special center, yet exhibit a clear-cut clustering that was not a deliberate input. They also exhibit "filaments" and "walls," which could not possibly have been imputed, because I did not know that they had been observed.

*Conjecture: could it be that the observed "clusters," "filaments" and "walls," need not be explained separately, but necessarily follow from "scale free" fractality?* This would mean that all those structures do not result from unidentified features of specific models but are unavoidable consequences of random fractality — as interpreted by a human brain.

The preceding paragraph is deficient insofar as the word "conjecture" cannot be given a strict mathematical meaning, unless a mathematical meaning is advanced for the remaining terms.

*Lacunarity.* A problem arose when careful examination of the simulations revealed a clearly incorrect prediction. The original *Seeded Universe* proved to be visually far more lacunar than the real world, in the sense mentioned in Section 4.3. This means that the holes are larger in the simulations than in reality. The *Parted Universe* model fared better, since its lacunarity can be adjusted at will and fitted to the actual distribution.

A lowered lacunarity is expressed by a positive correlation between masses in antipodal directions. Testing this specific conjecture is a challenge for those who analyze the data.

*Conjectured mathematical explanation of why one should expect the distribution of galaxies to be fractal.* In a cubic box in which opposite sides are identified to form a three-dimensional torus, consider a large array of point masses subjected to Newtonian attraction. The evolution of this array obeys the Laplace equation, with an essential novelty: the singularities of the solution — which are the positions of the points — are movable. The numerous simulations I know of (beginning with those performed at IBM around 1960) all suggest the following. Even when the pattern of the singularities begins by being uniform or Poisson, it gradually creates clusters and a semblance of hierarchy, and appears to tend toward fractality. It is

against the preceding background that I conjectured that the limit distribution of galaxies is fractal, and that the origin of fractality lies in Newton's equations.

#### 6.4 *The Navier-Stokes and Euler equations of fluid motion and the conjectured fractality of their singularities*

*Background.* The first concrete use of a Cantor dust in real spaces is found in a 1963 paper on noise records by Berger & Mandelbrot (reprinted in *SN*), a work near simultaneous with Kolmogorov's work on the intermittence of turbulence. After numerous experimental tests, designed to create an intuitive feeling for this phenomenon (e.g., listening to turbulent velocity records that were made audible), I extended the fractal viewpoint to turbulence, and was led circa 1964 to the following conjecture.

*Conjecture concerning the geometric nature of "turbulently dissipative" parts of spaces.* Dissipation should be viewed as occurring, not in domains in a fluid with significant interior points, but in fractal sets. In a first approximation, those sets' intersection with a straight line is a Cantor-like fractal dust having a dimension in the range from 0.5 to 0.6. The corresponding full sets in space should therefore be expected to be fractals with Hausdorff dimension in the range from 2.5 to 2.6.

Actually, Cantor dust and Hausdorff dimension are not the proper notions in the context of viscous fluids, because viscosity necessarily erases the fine detail that is essential to Cantor fractals. Hence the following weaker conjecture.

*Conjecture:* *FGN*, Chapter 11 and Mandelbrot, *C. R. Acad. Sc. Paris*: 282A, 1976, 119, translated as Chapter N19 of *SN*). The dissipation in a viscous fluid occurs in the neighborhood of a singularity of a nonviscous approximation following Euler's equations, and the motion of a nonviscous fluid acquires singularities that are sets of dimension about 2.5 to 2.6.

*Open mathematical problem:* To prove or disprove this conjecture, under suitable conditions.

*Comment A.* Several numerical tests agree with this conjecture (e.g., Chorin, *Commun. Pure and Applied Math.*: 34, 1981, 853).

*Comment B.* I also conjectured that the Navier-Stokes equations have fractal singularities of much smaller dimension. This conjecture has led to extensive work by V. Scheffer, R. Teman and C. Foias, and many others, but is not exhausted.

*Comment C.* As is well-known to students of chaos, a few years after my work, fractals in phase space entered the transition from laminar to turbulent flow, through the work of Ruelle and Takens and their followers. The task of unifying the roles of fractals in real and phase spaces is not completed.

## 7 Iterates of the complex map $z^2 + c$ . Julia and Mandelbrot sets

The study of iterates of rational functions of a complex variable is an old topic of pure mathematics that reached a sharp peak circa 1918 with Fatou and Julia. Those authors succeeded so well that — apart from the proof of the existence of Siegel discs — their theory remained largely unchanged for sixty years. A more recent sharp break began in 1980 and has become iconic since most "ordinary"

people seem to have heard of the Mandelbrot set: it is arguably the only tangible proof known to them that mathematics is alive and well. The beginnings are now fully documented in *SC* therefore a bare sketch will suffice here.

### 7.1 The $J$ -set or Julia set

The Julia set is defined as the repeller of rational iteration. For the quadratic map  $z \rightarrow z^2 + c$ , a more direct definition is available: the filled-in Julia set of a given  $c$  is the set of points that the map does not iterate to the point at infinity, and the Julia set is the boundary of the filled-in Julia set. With few exceptions, it is fractal: a nonanalytic curve or a "Cantor-like" dust. Julia called these sets "very irregular and complicated." The computer — which I was the first to use systematically — led to beautiful wildly colorful displays that must now be familiar to every reader. To associate forever the name of Fatou and Julia, the complement of the Julia set is best called the Fatou set and its maximal open components, Fatou domains.

Starting with the quadratic map, I explored numerically and graphically how the value of  $c$  affects the dynamics and the shape of the Julia set.

### 7.2 The set $M_0$ and the Mandelbrot set

*The  $M_0$  set.* Of greatest interest from the viewpoint of dynamics, hence of physics, is the set  $M_0$  of those values of  $c$  for which the map  $z \rightarrow z^2 + c$  has a finite stable limit cycle. This set having proved to be hard to investigate directly, I moved on to the computer-assisted investigation of a set that was easier to study and seemed closely related.

*The  $M$  set.* The set of those parameter values  $c$  in the complex plane, for which the Julia set is connected, was called the  $\mu$ -map in *FGN* (Chap. 19), but Douady and Hubbard called it *the Mandelbrot set*.

The Mandelbrot set proved to be a most worthy object of study, first for "experimental mathematics" and then for mathematics, and it also gave birth to a new form of art! It is so well and so widely known, that no further reference is needed. But it is good to mention that the  $M$  set is a universal object. Curry, Garnett, and Sullivan (*Commun. Math. Phys.*: 91, 1983, 267) discovered that  $M$  arises also in Newton's method for cubic polynomials, a dynamical system significantly different from  $z \rightarrow z^2 + c$ . Following this, Douady and Hubbard (*Ann. Sc. Ec. Norm. Sup. (Paris)*: 18, 1985, 287) developed the theory of quadratic-like maps and showed the  $M$  set arises for a wide variety of functions, and in this sense is a universal object.

Also, the study of  $z \rightarrow z^2 + c$  naturally suggested the study of similar questions for other polynomials. But even the generic cubic,  $z \rightarrow z^3 + az + b$ , has proved soberingly difficult. Intense study by extremely powerful mathematicians still leaves many questions unanswered.

### 7.3 Relations between $M_0$ and $M$ ; the incredibly stubborn conjecture that $M$ is the closure of $M_0$ ; "MLC"

Computer graphics approximates  $M_0$  by a smaller set and  $M$  by a larger set. Early on, extending the duration of the computation seemed to make the two represen-

tations converge to each other. Furthermore, when  $c$  is an interior point of  $M$ , not too close to the boundary, it was easily checked that a finite limit cycle exists. Those observations led me to conjecture that  $M$  is identical to  $M_0$  together with its limits points.

In terms of its being simple and understandable without any special preparation, this conjecture comes close to the “dimension  $4/3$ ” conjecture about Brownian motion, discussed in Section 2. Again, I could think of no proof, even of a heuristic one. More significantly, the conjecture remains unanswered.

*The MLC conjecture.* Many equivalent statements were identified, the best known being that the Mandelbrot set is locally connected. This statement was given a “nickname,” MLC. It has the great advantage of being local and was proven for a very large subset of the boundary of  $M$  — earning J. C. Yoccoz a Fields medal. But, compared to the original form, MLC has the great drawback of being far from intuitive. (For the generic cubic map, the corresponding local connectivity conjecture was proved to be false.)

## 8 Limit sets of Kleinian groups

A collection of Möbius transformations of the form  $z \rightarrow (az + b)/(cz + d)$  defines a group that Poincaré chose to call Kleinian. With few exceptions, their limit sets  $S$  are fractal. For the closely related groups based on geometric inversions in a collection  $C_1, C_2, \dots, C_n$  of circles, there is a well-known algorithm that yields  $S$  in the limit. But it converges with excruciating slowness as seen in the top panel of Figure 5. For a century, the challenge to obtain a fast algorithm remained unanswered, but I met it in many cases as seen in the middle panel of Figure 5. For details, see Chapter 18 of *FGN* and *Mathematical Intelligencer*: 5(2), 1983, 9, both reproduced in *SC*.

*An interesting contrast.* By leading to the  $4/3$  conjecture, fractal geometry opened a brand new mathematical problem and gave it a very active constituency; but it failed to contribute to solving it. With inversion groups, fractal geometry dealt with a *very old* problem long viewed as so difficult that it had long to have an active constituency. Not only was the problem solved to a significant degree, but it was made, in a literal sense, childishly easy: it is a nice example used in the high school classes examined in this paper’s Section 11.

*The fast algorithm first described in FGN and illustrated in Figure 6.* The limit set of the group of transformations generated by inversions covers the complement of  $S$  by a denumerable collection of circles that “osculate”  $S$ . The circles’ radii decrease rapidly, therefore their union outlines  $S$  very efficiently.

When  $S$  is a Jordan curve (as on Plate 177 of *FGN*), two collections of osculating circles outline  $S$ , respectively, from the inside and the outside. They are closely reminiscent of the collection of osculating triangles that outline Koch’s snowflake curve from both sides in a construction that is described in Plate 43 of *FGN* and dates to the 1900s. Because of this analogy, the osculating construction appears, after the fact, to be entirely “natural.” But this appearance is thoroughly misleading, as proven by the gap of roughly hundred years that elapsed before it was discovered. It was not obvious at all because of the mood of mathematics: even

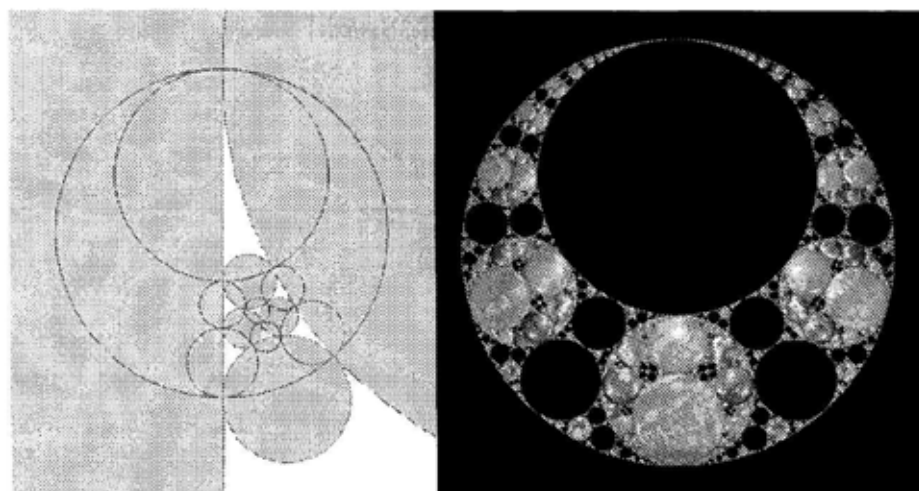


Figure 5. This is Figure C16-2 of *SC*. It elaborates upon Plate 199 of *FGN* and page 129 of *SN*. The “generator” part of the diagram consists in six circles filled in gray. The inversions with respect to those circles, when combined with prescribed probabilities, define a “decomposable dynamical system” also called IFS. The limit set is a self-inverse fractal for which I discovered a new algorithm using the diagram’s remaining eight bold circles. The decorative “Pharaoh Breastplate” represents four of those circles and their successive inverses represent by four kinds of “semi-precious stones.”

after computer graphics had become available, it continued to scorn pictures. The algorithm did not start to be viewed as natural until it literally burst out after respectful examination of pictures of many special examples.

A particularly striking example is seen in Figure 5, called “Pharaoh’s breast-plate,” a black-and-white rendering of Plate 199 of *FGN*, of the cover of *SN* and of a figure in *SC*. A more elaborate version of this picture appears on the cover of *SN*. This is the limit set of a group generated by inversion in the 6 circles drawn as thin lines on the small accompanying diagram. Here, the basic osculating circles actually belong to the limit set and do not intersect (each is the limit set of a Fuchsian subgroup based on three circles). The other osculating circles follow by all sequences of inversions in the 6 generators, meaning that each osculator generates a “clan” with its own color.

By inspection, it is easy to see this group also has three additional Fuchsian subgroups, each made of four generators and contributing full circles to the limit set.

Pictures such as Figure 5 are not only aesthetically pleasing, but they helped breathe new life into the study of Kleinian groups, recently exemplified by the book by Mumford, Series, & Wright: *Indra’s Pearls* (Cambridge University Press, 2002). Thurston’s work on hyperbolic geometry and 3-manifolds opens up the possibility for limit sets of Kleinian group actions to play a role in the attempts to classify 3-manifolds. The Hausdorff dimension of these limit sets has been studied by Sullivan, Canary and others.



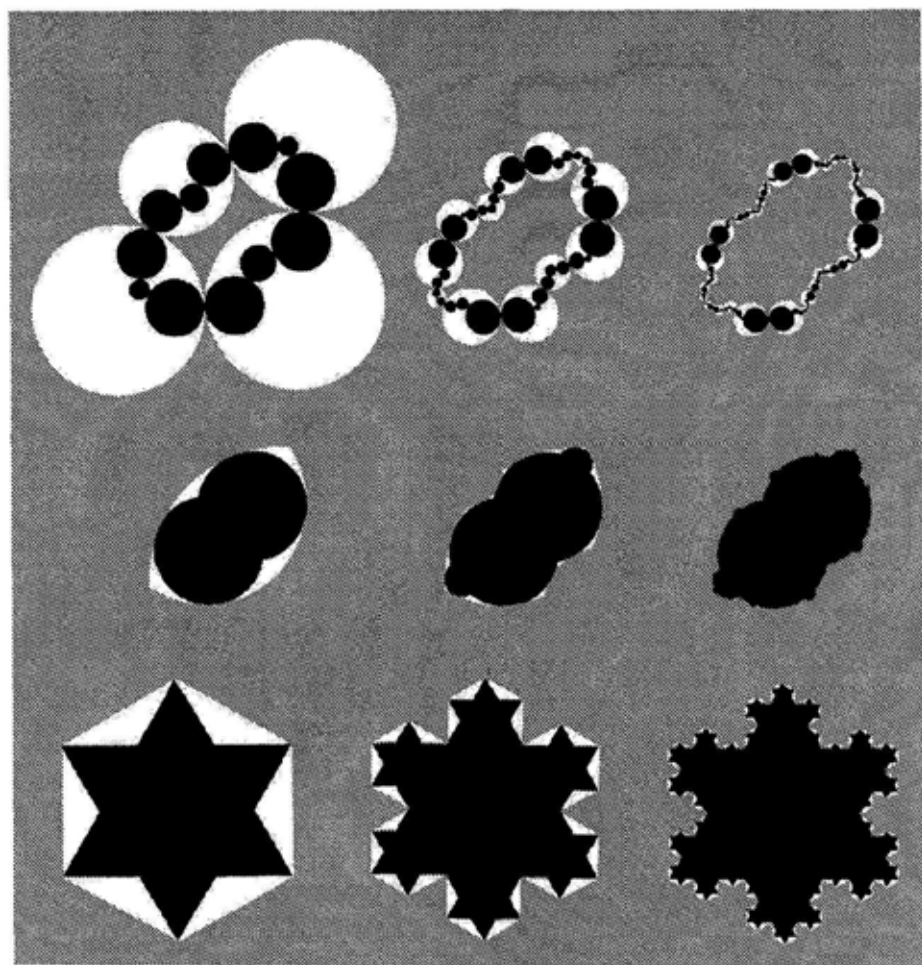


Figure 6. This is Figure C16-1 of *SC* and a composite of page 173 and Plates 177 and 43 of *FGN*. The two top panels represent two constructions of the limit set of a group based on inversions. The top panel shows the slowly converging classical construction (Poincaré). The middle panel shows my fast-converging proposed alternative. The latter recalls the Cesaro construction of the Koch snowflake that is shown in the lower panel.

*Challenge.* Incorporate lacunarity and multifractal measures into the study of 3-manifolds through these limit sets.



## 9 The study of power law probability distributions and the notion that variability and randomness can fall into distinct “states,” ranging from “mild” to “slow” and “wild”

### 9.1 *The evolution of power-law probability distributions, from a neglected periphery of statistics to a central position in fractal geometry and the topic of active interest on its own*

The most widely known analytic tool of fractal geometry consists in power-law relations and power-law probability distributions. They are ancient since Ohmori discovered such a law for earthquake aftershocks in 1894, predating even the Pareto law for the distribution of personal income which was discovered in 1896. Around 1950, however, power laws were viewed by statisticians and scientists alike as scattered anomalous. They were often arbitrarily replaced by the lognormal distribution, or otherwise questioned and played down. When I explained and demystified the Zipf law of word frequencies (*CR (Paris)* 232, 1951, 1638-1640 and 2003-2005), the situation changed completely. To bring power laws together credibly I devoted papers too numerous to be listed. References, reprints, and expositions are found in my *Selecta* books. In many sciences, those papers moved power laws to the forefront, interpreting them as evidence of the broad geometric scaling property of invariance that led to the concept of fractal.

### 9.2 *A basic distinction between “mild” and “wild” “states of variability:” practical aspects*

My early investigations of turbulence and price variation arose in the 1960s and used closely related procedures, thus confirming the saying that the Stock Market is as unpredictable and irregular as the weather. The analogy has gone much farther than one may have expected.

It led to general considerations about randomness that converged in Chapter 5 of *SE* to a distinction that may seem philosophical but is in fact very practical. In *principle*, Kolmogorov unified probability theory by providing unquestioned foundations. But in *practice* it is best to consider a function as belonging to one of several distinct “states of variability and/or randomness.” Among random variables, iconic examples are the Gaussian distribution for the *mild* state, the power-law distribution for the *wild* state, and the lognormal distribution for a *slow* state in between.

The key underlying fact contradicts a widespread but unfounded belief. The law of large numbers and the central limit theorem are *not* universal truths that one can blindly rely upon in model making. They are special properties that characterize cases of exceptional simplicity that define mild or slow randomness. The contrary is true of all the stochastic processes I used in investigating turbulence, finance and other fractal phenomena. All are examples of “wild” randomness.

This distinction deserves further discussion from the mathematical viewpoint and is bounded to play an active role in physics as well. The power-law long tails and/or dependence that rule all fractal phenomena are clear-cut symptoms of wild variability.

In many cases, the fractal or multifractal models that I put forward have been subjected to counter-claims. Alternative models put forward satisfy all the usual central limit theorems and appear to avoid both the formal mathematical difficulties and the “uncomfortable” consequences of wild randomness. Some of those models do not try hard and are content truncating the fractal models. Central limit behavior is thereby saved but only in an asymptotic sense that is useless because it is not reached in practice. Others proceed more indirectly but amount to the same thing.

*9.3 A small purely mathematical aspect of the mild-slow distinction. The boundary between these states provides the classical moment problem of classical analysis with a new wrinkle that originated (of all things!) in finance*

This subsection brings us back from finance to the purest mathematical analysis that flowered from stieltjes in the 1890s to the 1930s.

The boundary between the wild and slow states involves the classical central limit theorem, a key idea of probability theory. To the contrary, the boundary between the mild and the slow states is not at all traditional but marked by what I call the *criterion of short-run inequality*. Let  $P(x)$  be the tail probability of  $X$  and  $P_N(x)$ , the tail probability of the sum of  $N$  independent variables having the same distribution as  $X$ . Then, for fixed  $N$  and  $x$  tending to infinity, *FGN*, Chapter 5 showed the importance of the criterion that  $P_N(x)$  behaves like  $NP(x)$ .

That very simple criterion entered my work in 1960 for very practical reasons. But it turns out to run close to several complicated criteria that occur in the “moments problem” and the theory of quasi-analytic functions. This opened up a very interesting issue: could it be that time has come to study again those old topics once classical but lately very much out of fashion?

## 10 The variation of financial prices

Historically, my investigation of roughness was comparatively late in turning to physics and mathematics. It began in the early 1960s with investigations in economics that amounted to characterizing the roughness of financial charts. In the 1990s, this work became the foundation of “econophysics.” No other application illustrates more vividly the potency of the notion of fractal geometry as the beginning of a science of roughness.

In 1800, Louis Bachelier invented Brownian motion as a model of the variation of financial prices. Even before this model became widely accepted in academia, mine was the first voice to warn against its pitfalls. I pointed out that its two key features are thoroughly unrealistic, hence unacceptable. Having discovered that each involves an empirical power law distribution, I modelled both, first separately (Sections 10.1 and 10.2) and then jointly (Section 10.3), on the basis of the emerging concept of fractality. Under the term “scaling in finance,” this concept is the topic of Chapter 38 of *FGN*. Scaling became important in finance before it became important in the physics of criticality.

### 10.1 *The essential importance, even in a first approximation, of large sudden price discontinuities*

I was the first to argue that the neglect of discontinuities in the Brownian model is unjustified. They are not “outliers” one can safely disregard or study separately. To the contrary, I argued in 1963 (see *SE*) that their distribution is much more important than that of the “background noise” constituted by near Brownian small changes.

I followed this critique by showing that the big discontinuities and the small “noise” fall on a single power-law distribution and represented them by a scenario based on Levy stable distributions. Howard Taylor and I introduced in 1967 the new notion of intrinsic “trading time.” The originality of this work had been recognized all along. In 1964 P.H. Cootner called it “revolutionary.” Cootner also raised many questions that have all been answered. Forty years later, the “revolution” is bearing fruit in many diverse ways. Fractal trading time and my 1963 model have gained wide acceptance.

### 10.2 *The fact that the “background noise” of small price changes is of variable “volatility”*

That the so-called “price volatility” is itself “volatile” could not be denied but was ordinarily viewed as a symptom of non-stationality that must be studied separately. To the contrary (see *SE*), I interpreted this variability in 1965 as indicating that price changes differ from being statistically independent. In fact, for all practical purposes, their interdependence should be viewed as extending to an infinitely long term. Indeed, it too follows a power-law dependence. In particular, it is not limited to the short term that is studied by Markov processes and more recently ARCH or GARCH and its variants. I followed this critique and illustrated long dependence by introducing a process called fractional Brownian motion which has become very widely used.

### 10.3 *Multifractal models of price variation*

I introduced multifractality (minus the term) in 1968 in the context of turbulence (see *SN*). But I immediately observed and pointed out in 1972 (see *SE* and *SN*) that — because it combines long power-law tails and long power-law dependence — multifractality also apply to finance. I also introduced “cartoons” that realize long tails and long dependence and a very simple process understandable to experts and beginners alike.

*Fractional Brownian motion in multifractal time, and its use in financial modelling.* One half of *SE* is made of reprints of previously published works of mine, but Chapter 6 consists in material never previously published. It advances a new model of variation of prices that is further explored in many publications of mine, in particular, in *Quantitative Finance*: 1,2001, 113-123, 124-130, 427-440, 641-649, and 558-559.

This model represents price as a fractional Brownian motion  $B_H$ , that, instead of the clock time,  $t$ , is followed in a “trading time,”  $\theta$ . Those two times are related

by a multifractal function  $\theta(t)$  that is the integral of a random multifractal measure. That is,  $P(t) = B_H[\theta(t)]$ . At this early stage of the theory, I assumed the functions  $B_H$  and  $\theta$  to be statistically independent. This process is specified by the properties of  $B_H$ , primarily its exponent  $H$  (a Hölder exponent) and the properties of  $\theta(t)$ , beginning with its  $f(\alpha)$  spectrum. This process was found to fit diverse financial data very well. From most other viewpoints, it is wide open for exploration.

## 11 The directly useful fractal

Early on, I used to point out a striking contrast: in raw nature smooth shapes are rare exceptions but in manufactured goods they were the near-universal rule. Tables are meant to be horizontal planes with near-linear or circular edges. Walls are meant to be vertical planes.

My early standard of fractality, the Eiffel Tower, was not accepted as counterexample: it remains a masterpiece of engineering but one never meant to be useful. Engineering seemed to be a systematic reaction against the roughness of raw nature.

An invidious claim added to those voiced in Sections 1.4 and 1.6 was that fractals have not contributed to any existing engineering problems.

All this initially led to a question "Did I expect fractals to become practically useful, and, if so, how soon?" I used to recommend patience, recalling the fate of astronomy: while every stage in its development had immediate users who helped support it, all those users were astrologers.

In due time (and with no direct help from me) fractals have indeed become widely useful. Too bad that each real use hits only a specific group of users, so that hardly anyone notices. The following list, very schematic and incomplete, can only touch fields that allow open publication. This excludes finance where what is published may never reflect what is actually used.

*Traffic on the internet.* Early efforts to squeeze the traffic's extreme variability into the familiar Poisson process soon failed. The multifractal model is now generally acknowledged as being the best and it is the topic of intense study.

*Road traffic.* The data are less abundant but one often needs multifractality.

*Antennas.* Stick antennas' properties are easy to analyze mathematically but inadequate and for antennas made of even a few sticks the mathematical analysis rapidly becomes very complicated. The properties of fractal antennas are both far better and easily calculated.

*Capacitors.* To achieve one farad, flat capacitors need a very large area and a little folding makes the mathematical analysis very complicated. Fractally folded one-farad capacitors are easy to calculate and fit on a pinhead.

*Sound-absorbing road barriers.* Houses close to roads want to be protected from traffic noise. Early on, flat protection panels simply reflected noise. Incoming panels with a fractal pattern are noise-absorbent.

*Chemical engineering.* When two gases are meant to react, it is best to control the surface of reaction. This is achieved by bringing one reactant in the midst of the other with the help of a spatial tree. The reaction is faster and cleaner with fewer impurities.

## 12 Fractals in the college and school classroom

Several examples in this paper share a very nice feature that is also very unusual. Among fields of research, fractal geometry may well exhibit some of the shortest distances and the greatest contrasts between a straightforward core and multiple new frontiers. The latter are filled with major difficulties of every kind, including conjectures that everybody can understand but no one can prove.

Starting with *FGN*, the core has by now become widely known, even to children and adult amateurs. This has opened a wonderful new opportunity that deserves brief mention all by itself.

At issue is the abyss between mathematics and a wider community. Its story is old but in the 1960s and 1970s the “new math” fad made it deeper. I think that no one benefits from this abyss, yet some continue to welcome it, and many more can think of no suitable bridge and view the abyss as inevitable. Therefore, it persists. M.L. Frame and I have convinced the Mathematics Department at Yale University that, in fact, a strong bridge can be based on fractal geometry.

The upshot: for the last several years, Yale has been offering an undergraduate course and associated summer workshops that teach fractals to several groups of non-mathematicians. Their attractiveness to students depends heavily on three assets.

*One* is the already mentioned unusually short distance from the simple to the complex and even the impossibly difficult. To the contrary, from the viewpoint of mathematics education, one of the worst features of most topics is that prerequisites are interminable. They are unavoidable but respond to needs that do not become compelling until the ends of long paths that allow many opportunities to drop out.

*A second asset* is that the history of fractals reaches back for several millennia, proving that fractality is “natural” to the culture of our species.

*A third asset* is, of course, that the ubiquity of roughness translates into a large number and variety of current applications of every kind in the works of Nature and Man.

*Fourth asset:* as a very valuable by-product, our course teaches the meaning of rigor by the most efficient method: when a program is buggy, the computer immediately screams Error! at the programmer.

The book *FM* explains and motivates our course and reproduces stories from several colleagues who work along the same lines. It also refers to two items on the web: an extensive set *course-notes* and a *Panorama* that collects innumerable examples of fractality. Everyone is invited to add to this collection!

## Acknowledgements

Conversations with Professor Michael Frame of Yale University helped greatly in the preparation of this text. Professor Kenneth Monks of the University of Scranton drew Figure 3 using the algorithm described in *FGN*.

## References

### Style of reference and books referenced by italic capitals

References to serials are scattered through the paper at the proper places. The books are denoted by letters suggested by various mnemonic devices: initials of the author or the titles or (in the case of *Selecta*) the initials of economics, noise, Hurst, and chaos.

- FM Frame, M. and Mandelbrot, B.B. 2002, *Fractals, Graphics and Mathematics Education*. Mathematical Association of America and Cambridge University Press.
- FGN Mandelbrot, B. B. 1982, *The Fractal Geometry of Nature*, W. H. Freeman and Co., New York and Oxford. The second and later printings include an Update and additional references. Earlier versions were *Les objets fractals: forme, hasard et dimension*, Flammarion, Paris, 1975 (fourth edition, 1995) and *Fractals: Form, Chance and Dimension*, Freeman, 1977. There are innumerable translations, for example, the 1975 book was translated into Basque.
- SE Mandelbrot, B. B. 1997E, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (Selecta, Volume E) Springer-Verlag, New York.
- SN Mandelbrot, B. B. 1999N, *Multifractality and 1/f Noise: Wild Self-Affinity in Physics*. (Selecta, Volume N). Springer-Verlag, New York.
- SH Mandelbrot, B. B. 2002H, *Gaussian Self-Affinity and Fractals: Global Dependence, R/S, 1/f, Rivers & Reliefs*. (Selecta, Volume H). Springer-Verlag, New York.
- SC Mandelbrot, B. B. 2004C, *Fractals and Chaos: the Mandelbrot Set and Beyond*. (Selecta, Volume C). Springer-Verlag, New York.
- ST Mandelbrot, B. B. 2004T, *Thermometry and Thermodynamics: Foundations and Generalization*. Webbook.

This page intentionally left blank

# A RENEWAL PROCESS OF MITTAG-LEFFLER TYPE

FRANCESCO MAINARDI

*Dipartimento di Fisica, Università di Bologna and INFN,  
Via Irnerio 46, I-40126 Bologna, Italy  
E-mail: mainardi@bo.infn.it URL: www.fracalmo.org*

RUDOLF GORENFLO

*Department of Mathematics and Informatics, Free University of Berlin,  
Arnimallee 3, D-14195 Berlin, Germany  
E-mail: gorenflo@mi.fu-berlin.de*

ENRICO SCALAS

*Dipartimento di Scienze e Tecnologie Avanzate, Università del  
Piemonte Orientale, via Cavour 84, I-15100 Alessandria, Italy  
E-mail: scalas@unipmn.it*

The purpose of this paper is to provide a generalization of the Poisson renewal process and the related Erlang distribution via fractional calculus. We first recall the basic renewal theory including its fundamental concepts like waiting time between events, the counting function and its average, the survival probability. If the waiting time is exponentially distributed we have a Poisson process, which is Markovian. However, other waiting time distributions are also relevant in applications, in particular those with a fat tail caused by a power law decay of its density. In this context we analyze a non-Markovian renewal process with a waiting time distribution described by the Mittag-Leffler function. This distribution, containing the exponential one as limiting case, is shown to play a fundamental role in the infinite thinning procedure of a generic renewal process governed by a power-asymptotic waiting time.

*MSC 2000:* 26A33, 33B20, 33E12, 44A10, 44A35, 60G55, 60J05, 60K05.

## 1 Essentials of renewal theory

The concept of *renewal process* has been developed as a stochastic model for describing the class of counting processes for which the times between successive events are independent identically distributed (*iid*) non-negative random variables, distributed with an arbitrary distribution. These times are referred to as waiting times or inter-arrival times.

The renewal processes generalize the classical Poisson process, that is known to be a counting process where the waiting times are exponentially distributed. For an example of a renewal process, suppose that we have an infinite supply of light-bulbs whose lifetimes are independent, identically distributed. Suppose also that we use a single light-bulb at a time, and when it fails we immediately replace it with a new one. Under these conditions,  $\{N(t), t \geq 0\}$  is a renewal process when  $N(t)$  represents the number of light-bulbs that have failed. For more details see *e.g.* the classical treatises by Khintchine <sup>10</sup>, Cox <sup>3</sup>, Gnedenko & Kovalenko <sup>6</sup>, Feller <sup>5</sup>, and the recent book by Ross <sup>15</sup>.



For a renewal process having waiting times  $T_1, T_2, \dots$ , let

$$t_0 = 0, \quad t_k = \sum_{j=1}^k T_j, \quad k \geq 1. \quad (1.1)$$

That is  $t_1 = T_1$  is the time of the first renewal,  $t_2 = T_1 + T_2$  is the time of the second renewal and so on. In general  $t_k$  denotes the  $k$ th renewal.

The process is specified if we know the probability law for the waiting times. In this respect we introduce the *probability density function* (pdf)  $\phi(t)$  and the (cumulative) distribution function  $\Phi(t)$  so defined:

$$\phi(t) := \frac{d}{dt} \Phi(t), \quad \Phi(t) := P(T \leq t) = \int_0^t \phi(t') dt'. \quad (1.2)$$

In the above example, where the nonnegative random variable represent the lifetime of technical systems, one refers to  $\Phi(t)$  as to the *failure probability* and to

$$\Psi(t) := P(T > t) = \int_t^\infty \phi(t') dt' = 1 - \Phi(t) \quad (1.3)$$

as to the *survival probability*, because  $\Phi(t)$  and  $\Psi(t)$  are the respective probabilities that the system does or does not fail in  $(0, T]$ . These terms, however, are commonly adopted for any renewal process.

A relevant quantity is the *counting function*  $N(t)$  defined as

$$N(t) := \max \{k | t_k \leq t, k = 0, 1, 2, \dots\}, \quad (1.4)$$

that represents the effective number of events before or at instant  $t$ . In particular we have  $\Psi(t) = P(N(t) = 0)$ . Continuing in the general theory we set  $F_1(t) = \Phi(t)$ ,  $f_1(t) = \phi(t)$ , and in general

$$F_k(t) := P(t_k = T_1 + \dots + T_k \leq t), \quad f_k(t) = \frac{d}{dt} F_k(t), \quad k \geq 1, \quad (1.5)$$

thus  $F_k(t)$  represents the probability that the sum of the first  $k$  waiting times is less or equal  $t$  and  $f_k(t)$  its density. Then, for any fixed  $k \geq 1$  the normalization condition for  $F_k(t)$  is fulfilled because

$$\lim_{t \rightarrow \infty} F_k(t) = P(t_k = T_1 + \dots + T_k < \infty) = 1. \quad (1.6)$$

In fact, the sum of  $k$  random variables each of which is finite with probability 1 is finite with probability 1 itself. By setting for consistency  $F_0(t) \equiv 1$  and  $f_0(t) = \delta(t)$ , the Dirac delta function<sup>a</sup>, we also note that for  $k \geq 0$  we have

$$P(N(t) = k) := P(t_k \leq t, t_{k+1} > t) = \int_0^t f_k(t') \Psi(t - t') dt'. \quad (1.7)$$

<sup>a</sup>We find it convenient to recall the *formal representation* of this generalized function in  $\mathbf{R}^+$ ,

$$\delta(t) := \frac{t^{-1}}{\Gamma(0)}, \quad t \geq 0.$$

A related quantity is the *renewal function*  $m(t)$  defined as

$$m(t) := E(N(t)) = \langle N(t) \rangle = \sum_{k=1}^{\infty} P(t_k \leq t), \quad (1.8)$$

that represents the average number of events in the interval  $(0, t]$ . The renewal function uniquely determines the renewal process, see *e.g.* <sup>15</sup>. It turns out to be related to the waiting time distribution by the so-called *Renewal Equation*,

$$m(t) = \Phi(t) + \int_0^t m(t-t') \phi(t') dt' = \int_0^t [1 + m(t-t')] \phi(t') dt'. \quad (1.9)$$

If the mean waiting time (the first moment) is finite, namely

$$\mu := \langle T \rangle = \int_0^{\infty} t \phi(t) dt < \infty, \quad (1.10)$$

it is known that, with probability 1,  $t_k/k \rightarrow \mu$  as  $k \rightarrow \infty$ , and  $N(t)/t \rightarrow 1/\mu$  as  $t \rightarrow \infty$ , which imply the *Elementary Renewal Theorem*,

$$\frac{m(t)}{t} \rightarrow \frac{1}{\mu} \quad \text{as } t \rightarrow \infty. \quad (1.11)$$

However, when the waiting time laws exhibit fat tails, the mean waiting time may be infinite as in the case of the power law asymptotics considered later,

$$\phi(t) = O(t^{-(1+\beta)}) \quad \text{as } t \rightarrow \infty, \quad \text{if } 0 < \beta < 1. \quad (1.12)$$

We now find it convenient to introduce the simplified  $*$  notation for the Laplace convolution between two causal well-behaved (generalized) functions  $f(t)$ ,  $g(t)$

$$\int_0^t f(t') g(t-t') dt' = f(t) * g(t) = g(t) * f(t) = \int_0^t f(t-t') g(t') dt'.$$

Since  $f_k(t)$  is the *pdf* of the sum of the  $k$  *iid* random variables  $T_1, \dots, T_k$  with *pdf*  $\phi(t)$ , and  $f_0(t) = \delta(t)$ , we easily recognize that

$$f_k(t) = [\phi(t)]^{*k}, \quad k = 0, 1, 2, \dots, \quad (1.13)$$

so Eq. (1.7) simply reads:

$$P(N(t) = k) = [\phi(t)]^{*k} * \Psi(t), \quad k = 0, 1, 2, \dots. \quad (1.14)$$

We also note that in this notation the *renewal equation* (1.9) reads

$$m(t) = \Phi(t) + m(t) * \phi(t). \quad (1.15)$$

Because of the presence of Laplace convolutions a renewal process is suited for the Laplace transform method. Throughout this paper we will denote by  $\tilde{f}(s)$  the Laplace transform of a sufficiently well-behaved (generalized) function  $f(t)$  according to

$$\mathcal{L}\{f(t); s\} = \tilde{f}(s) = \int_0^{+\infty} e^{-st} f(t) dt, \quad s > s_0,$$

and for  $\delta(t)$  consistently we will have  $\tilde{\delta}(s) \equiv 1$ . Note that for our purposes we agree to take  $s$  real. Then, in the Laplace domain, the *renewal equation* reads

$$\tilde{m}(s) = \tilde{\Phi}(s) + \tilde{m}(s) \tilde{\phi}(s), \quad (1.16)$$

with  $\tilde{\Phi}(s) = \tilde{\phi}(s)/s$ , from which

$$\tilde{m}(s) = \frac{\tilde{\phi}(s)}{s [1 - \tilde{\phi}(s)]}, \quad \tilde{\phi}(s) = \frac{\tilde{s} m(s)}{1 + s \tilde{m}(s)}. \quad (1.17)$$

In view of (1.13) we recognize that Eq. (1.7) reads in the Laplace domain

$$\mathcal{L}\{P(N(t) = k); s\} = [\tilde{\phi}(s)]^k \tilde{\Psi}(s), \quad k = 0, 1, 2, \dots, \quad (1.18)$$

where, using (1.3),

$$\tilde{\Psi}(s) = \frac{1 - \tilde{\phi}(s)}{s}. \quad (1.19)$$

## 2 The Poisson process as a renewal process

The most celebrated renewal process is the Poisson process characterized by a waiting time *pdf* of exponential type,

$$\phi(t) = \lambda e^{-\lambda t}, \quad \lambda > 0, \quad t \geq 0. \quad (2.1)$$

Then the moments turn out to be

$$\langle T \rangle = \frac{1}{\lambda}, \quad \langle T^2 \rangle = \frac{1}{\lambda^2}, \dots, \langle T^n \rangle = \frac{1}{\lambda^n}, \dots \quad (2.2)$$

and the *survival probability*

$$\Psi(t) := P(T > t) = e^{-\lambda t}, \quad t \geq 0. \quad (2.3)$$

The exponential distribution is characteristic of a process without memory. We know that the probability that  $k$  events occur in the interval of length  $t$  is just

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geq 0. \quad (2.4)$$

Furthermore the Poisson process turns out to be a renewal process with linear renewal function, namely

$$m(t) = \lambda t, \quad t \geq 0. \quad (2.5)$$

The probability distribution related to the sum of  $k$  *iid* exponential random variables is known to be the so-called *Erlang distribution* (of order  $k$ ). The corresponding density (the *Erlang pdf*) is thus

$$f_k(t) = \lambda \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t}, \quad t \geq 0, \quad k \geq 1, \quad (2.6)$$

so that the Erlang distribution function turns out to be

$$F_k(t) = \int_0^t f_k(t') dt' = 1 - \sum_{n=0}^{k-1} \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=k}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad t \geq 0. \quad (2.7)$$

In the limiting case  $k = 0$  we recover  $f_0(t) = \delta(t)$ ,  $F_0(t) \equiv 1$ ,  $t \geq 0$ .

The results (2.4)-(2.7) can easily be obtained by using the technique of the Laplace transform sketched in the previous section noting that for the Poisson process we have:

$$\tilde{\phi}(s) = \frac{\lambda}{\lambda + s}, \quad \tilde{\Psi}(s) = \frac{1}{\lambda + s}, \quad \tilde{m}(s) = \frac{\lambda}{s^2}, \quad (2.8)$$

and for the Erlang distribution

$$\tilde{f}_k(s) = [\tilde{\phi}(s)]^k = \frac{\lambda^k}{(\lambda + s)^k}, \quad \tilde{F}_k(s) = \frac{[\tilde{\phi}(s)]^k}{s} = \frac{\lambda^k}{s(\lambda + s)^k}. \quad (2.9)$$

We also recall that the survival probability for the Poisson renewal process obeys the ordinary differential equation (of relaxation type)

$$\frac{d}{dt} \Psi(t) = -\lambda \Psi(t), \quad t > 0; \quad \Psi(0^+) = 1. \quad (2.10)$$

### 3 A fractional generalization of the renewal Poisson process

A fractional generalization of the renewal Poisson process is simply obtained by generalizing the differential equation (2.10) replacing there the first derivative with the integro-differential operator  ${}_t D_*^\beta$  that is interpreted as the fractional derivative of order  $\beta$  in Caputo's sense, see Appendix A. We write, taking for simplicity  $\lambda = 1$ ,

$${}_t D_*^\beta \Psi(t) = -\Psi(t), \quad t > 0, \quad 0 < \beta \leq 1; \quad \Psi(0^+) = 1. \quad (3.1)$$

We also allow the limiting case  $\beta = 1$  where all the results of the previous section (with  $\lambda = 1$ ) are expected to be recovered.

For this purpose we have to recall the Mittag-Leffler function as the natural "fractional" generalization of the exponential function, that characterizes the Poisson process. The Mittag-Leffler function of parameter  $\beta$ , is defined in the complex plane by the power series

$$E_\beta(z) := \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\beta n + 1)}, \quad \beta > 0, \quad z \in \mathbb{C}. \quad (3.2)$$

It turns out to be an entire function of order  $\beta$  which reduces for  $\beta = 1$  to  $\exp(z)$ . For detailed information on the Mittag-Leffler-type functions and their Laplace transforms the reader may consult e.g. <sup>4,8,12,14</sup>.

The solution of Eq. (3.1) is known to be, see e.g. <sup>2,8,11</sup>,

$$\Psi(t) = E_\beta(-t^\beta), \quad t \geq 0, \quad 0 < \beta \leq 1, \quad (3.3)$$

so

$$\phi(t) := -\frac{d}{dt} \Psi(t) = -\frac{d}{dt} E_\beta(-t^\beta), \quad t \geq 0, \quad 0 < \beta \leq 1, \quad (3.4)$$

Then the corresponding Laplace transforms read:

$$\tilde{\Psi}(s) = \frac{s^{\beta-1}}{1+s^\beta}, \quad \tilde{\phi}(s) = \frac{1}{1+s^\beta}, \quad 0 < \beta \leq 1. \quad (3.5)$$

The renewal function can be deduced from its Laplace transform by using (1.17) and (3.5); we find

$$\tilde{m}(s) = \frac{1}{s^{1+\beta}}; \quad m(t) = \frac{t^\beta}{\Gamma(1+\beta)}, \quad t \geq 0, \quad 0 < \beta \leq 1. \quad (3.6)$$

For  $\beta < 1$  it turns out super-linear for small  $t$  and sub-linear for large  $t$ .

Hereafter, we find it convenient to summarize the most relevant features of the functions  $\Psi(t)$  and  $\phi(t)$  when  $0 < \beta < 1$ . We begin to quote their series expansions for  $t \geq 0$  and asymptotics for  $t \rightarrow \infty$ ,

$$\Psi(t) = \sum_{n=0}^{\infty} (-1)^n \frac{t^{\beta n}}{\Gamma(\beta n + 1)} \sim \frac{\sin(\beta\pi)}{\pi} \frac{\Gamma(\beta)}{t^\beta}, \quad (3.7)$$

and

$$\phi(t) = \frac{1}{t^{1-\beta}} \sum_{n=0}^{\infty} (-1)^n \frac{t^{\beta n}}{\Gamma(\beta n + \beta)} \sim \frac{\sin(\beta\pi)}{\pi} \frac{\Gamma(\beta+1)}{t^{\beta+1}}. \quad (3.8)$$

In contrast to the Poissonian case  $\beta = 1$ , in the case  $0 < \beta < 1$  for large  $t$  the functions  $\Psi(t)$  and  $\phi(t)$  no longer decay exponentially but algebraically. As a consequence of (3.8) we find the power-law asymptotics (1.9) for the waiting time *pdf* and the process turns out to no longer be Markovian but of long-memory type. However, we recognize that for  $0 < \beta < 1$  both functions  $\Psi(t)$ ,  $\phi(t)$  keep the "completely monotonic" character of the Poissonian case. Complete monotonicity of the functions  $\Psi(t)$  and  $\phi(t)$  means

$$(-1)^n \frac{d^n}{dt^n} \Psi(t) \geq 0, \quad (-1)^n \frac{d^n}{dt^n} \phi(t) \geq 0, \quad n = 0, 1, 2, \dots, t \geq 0, \quad (3.9)$$

or equivalently, their representability as (real) Laplace transforms of non-negative functions, see *e.g.* <sup>8</sup>.

To point out the behaviour of the Mittag-Leffler functions, in Figure 1 we exhibit plots of the survival probability  $\Psi(t) = E_\beta(-t^\beta)$  versus  $t$  for some values of the parameter  $\beta$ .

For the generalizations of Eqs (2.4) and (2.6)-(2.7), characteristic of the Poisson and Erlang distributions respectively, we must point out the Laplace transform formula

$$\mathcal{L}\{t^{\beta k} E_\beta^{(k)}(-t^\beta); s\} = \frac{k! s^{\beta-1}}{(1+s^\beta)^{k+1}}, \quad \beta > 0, \quad k = 0, 1, 2, \dots \quad (3.10)$$

with  $E_\beta^{(k)}(z) := \frac{d^k}{dz^k} E_\beta(z)$ , that can be deduced from the book by Podlubny, see (1.80) in <sup>14</sup>. For reader's convenience we report in Appendix B our proof of (3.10) adapted from <sup>14</sup>.

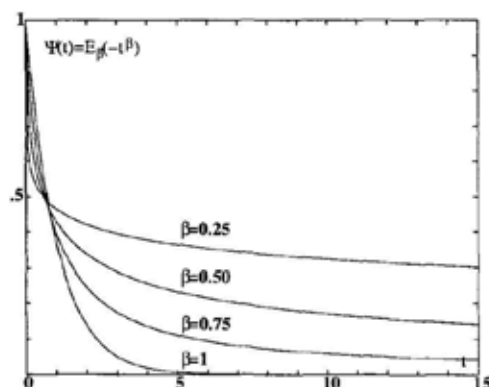


Figure 1. The function  $\Psi(t) = E_{\beta}(-t^{\beta})$  versus  $t$  for  $\beta = 0.25, 0.50, 0.75, 1$ . For  $\beta < 1$  we note the fast (algebraic) decay for  $t \rightarrow 0$  and the slow (algebraic) decay for  $t \rightarrow \infty$ ; for  $\beta = 1$  we recover the exponential function  $\exp(-t)$ .

Then, by using the Laplace transform pairs (3.5) and Eqs (3.3), (3.4), (3.10) in Eqs (1.13)-(1.14), we have the *generalized Poisson distribution*,

$$P(N(t) = k) = \frac{t^{k\beta}}{k!} E_{\beta}^{(k)}(-t^{\beta}), \quad k = 1, 2, \dots \quad (3.11)$$

and the *generalized Erlang pdf* (of order  $k \geq 1$ ),

$$f_k(t) = \beta \frac{t^{k\beta-1}}{(k-1)!} E_{\beta}^{(k)}(-t^{\beta}). \quad (3.12)$$

The corresponding *generalized Erlang distribution* turns out to be

$$F_k(t) = \int_0^t f_k(t') dt' = 1 - \sum_{n=0}^{k-1} \frac{t^{n\beta}}{n!} E_{\beta}^{(n)}(-t^{\beta}) = \sum_{n=k}^{\infty} \frac{t^{n\beta}}{n!} E_{\beta}^{(n)}(-t^{\beta}). \quad (3.13)$$

It is instructive to consider the special case  $\beta = 1/2$  for which it is known that

$$E_{1/2}(-\sqrt{t}) = e^t \operatorname{erfc}(\sqrt{t}) = e^t \frac{2}{\sqrt{\pi}} \int_{\sqrt{t}}^{\infty} e^{-u^2} du, \quad t \geq 0, \quad (3.14)$$

where  $\operatorname{erfc}$  denotes the *complementary error function*. In this case we can take profit of the recurrence relations for repeated integrals of the error functions, see e.g. <sup>1</sup>, §7.2, pp 299-300, to compute the derivatives of the Mittag-Leffler functions. For this purpose we recall:

$$\frac{d^n}{dz^n} (e^{z^2} \operatorname{erfc}(z)) = (-1)^n 2^n n! e^{z^2} I^n \operatorname{erfc}(z), \quad (3.15)$$

$$I^n \operatorname{erfc}(z) = \int_z^{\infty} I^{n-1} \operatorname{erfc}(\zeta) d\zeta, \quad n = 0, 1, 2, \dots, \quad (3.16)$$

$$I^{-1} \operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2}, \quad I^0 \operatorname{erfc}(z) = \operatorname{erfc}(z). \quad (3.17)$$

#### 4 The Mittag-Leffler distribution as limit for thinned renewal processes

The procedure of thinning (or rarefaction) for a generic renewal process (characterized by a generic random sequence of waiting times  $\{T_k\}$ ) has been considered and investigated by Gnedenko and Kovalenko<sup>6</sup>. It means that for each positive index  $k$  a decision is made: the event is deleted with probability  $p$  or it is maintained with probability  $q = 1 - p$ , with  $0 < q < 1$ . For this thinned or rarefied renewal process we shall hereafter revisit and complement the results available in<sup>6</sup>. We begin to rescale the time variable  $t$  replacing it by  $t/r$ , with a parameter  $r$  on which we will dispose later. Denoting, like in (1.5), by  $F_k(t)$  the probability distribution function of the sum of  $k$  waiting times and by  $f_k(t)$  its density, we have recursively, in view of (1.13),

$$f_1(t) = \phi(t), \quad f_k(t) = \int_0^t f_{k-1}(t-t') \phi(t') dt' = [\phi(t)]^{*k}, \quad k \geq 2. \quad (4.1)$$

Let us denote by  $(T_{q,r}f)(t)$  the waiting time density in the thinned and rescaled process from one event to the next. Observing that after a maintained event the next one of the original process is kept with probability  $q$  but dropped in favour of the second next with probability  $pq$  and, generally,  $n-1$  events are dropped in favour of the  $n$ -th next with probability  $p^{n-1}q$ , we arrive at the formula

$$(T_{q,r}f)(t) = \sum_{n=1}^{\infty} q p^{n-1} f_n(t/r)/r. \quad (4.2)$$

Let  $\tilde{f}_n(s) = \int_0^{\infty} e^{-st} f_n(t) dt$  be the Laplace transform of  $f_n(t)$ . Recalling  $f_1(t) = \phi(t)$  we set  $\tilde{f}_1(s) = \tilde{\phi}(s)$ . Then  $f_n(t/r)/r$  has the transform  $\tilde{f}_n(rs) = (\tilde{\phi}(rs))^n$ , and we obtain (in view of  $p = 1 - q$ ) the formula

$$(T_{q,r}\tilde{\phi})(s) = \sum_{n=1}^{\infty} q p^{n-1} (\tilde{\phi}(rs))^n = \frac{q \tilde{\phi}(rs)}{1 - (1-q)\tilde{\phi}(rs)}, \quad (4.3)$$

from which by Laplace inversion we can, in principle, construct the transformed process.

We now imagine stronger and stronger rarefaction (infinite thinning) by considering a scale of processes with the parameters  $r = \delta$  and  $q = \epsilon$  tending to zero under a scaling relation  $\epsilon = \epsilon(\delta)$  yet to be specified. Gnedenko and Kovalenko have, among other things, shown that if the condition

$$\tilde{\phi}(s) = 1 - a(s) s^{\beta} + o(a(s) s^{\beta}), \quad \text{for } s \rightarrow 0^+, \quad (4.4)$$

where  $a(s)$  is a slowly varying function<sup>b</sup> for  $s \rightarrow 0$ , is satisfied, then we have with

<sup>b</sup>**Definition:** We call a (measurable) positive function  $a(y)$ , defined in a right neighbourhood of zero, *slowly varying at zero* if  $a(cy)/a(y) \rightarrow 1$  with  $y \rightarrow 0$  for every  $c > 0$ . We call a (measurable) positive function  $b(y)$ , defined in a neighbourhood of infinity, *slowly varying at infinity* if  $b(cy)/b(y) \rightarrow 1$  with  $y \rightarrow \infty$  for every  $c > 0$ . Examples:  $(\log y)^{\gamma}$  with  $\gamma \in \mathbf{R}$  and  $\exp(\log y / \log \log y)$ .

$\epsilon = \epsilon(\delta) = a(\delta) \delta^\beta$  for every fixed  $s > 0$  the limit relation

$$\tilde{\phi}_0(s) := \lim_{\delta \rightarrow 0} \frac{\epsilon(\delta) \tilde{\phi}(\delta s)}{1 - (1 - \epsilon) \tilde{\phi}(\delta s)} = \frac{1}{1 + s^\beta}, \quad 0 < \beta \leq 1. \quad (4.5)$$

This condition is met with  $a(s) = \lambda M(1/s)$  if the waiting time  $T$  obeys a power law with index  $\beta$ , in the sense of *Master Lemma 2* by Gorenflo and Abdel-Rehim<sup>7</sup>. The function  $M(y)$  is the same as in *Master Lemma 2*, so it varies slowly at infinity, whence  $M(1/s)$  varies slowly at zero. The proof of (4.5) is by straightforward calculation. Observe the slow variation property of  $a(s)$  and note that terms small of higher order become negligible in the limit. By the continuity theorem for Laplace transforms, see Feller<sup>5</sup>, we now recognize  $\phi_0(t)$  as the limiting density, which we identify, in view of (3.2)-(3.5),

$$\phi_0(t) = -\frac{d}{dt} E_\beta(-t^\beta). \quad (4.6)$$

So the limiting waiting time density is the so-called Mittag-Leffler density, that in the special case  $\beta = 1$  reduces to the well-known exponential density,  $\exp(-t)$ . It should be noted that Gnedenko and Kovalenko in the sixties failed to recognize  $\tilde{\phi}_0(s)$  as Laplace transform of a Mittag-Leffler type function<sup>c</sup>.

## 5 Conclusions

We have provided a detailed analysis of the *fractional generalization* of the Poisson renewal processes by replacing the first time derivative in the relaxation equation of the survival probability by a fractional derivative of order  $\beta$  ( $0 < \beta \leq 1$ ). Consequently, we have obtained for  $0 < \beta < 1$  non-Markovian renewal processes where, essentially, the exponential probability densities, typical for the Poisson processes, are replaced by functions of Mittag-Leffler type, that decay in a power law manner with an exponent related to  $\beta$ . The renewal function of these processes is no longer proportional to time but to a power of time with exponent  $\beta$ .

The distributions obtained by considering the sum of  $k$  iid random variables ( $k = 1, 2, \dots$ ), distributed according to the Mittag-Leffler law provide the “fractional” generalization of the corresponding Erlang distributions (of order  $k$ ). Furthermore, the Mittag-Leffler probability distribution is shown to be the limiting distribution for the thinning procedure of a generic renewal process with waiting time density of power law character.

These results are useful to treat renewal processes *with reward*, so providing the “fractional” generalization of the compound Poisson processes, in the physical framework of uncoupled continuous-time random walks (i.e. random walks subordinated to a renewal process). In such processes, occurring in time and in space,

<sup>c</sup>Although the Mittag-Leffler function was introduced by the Swedish mathematician G. Mittag-Leffler in the first years of the twentieth century, it lived for long time in isolation as *Cinderella*. The term *Cinderella function* was used in the fifties by the Italian mathematician F.G. Tricomi for the incomplete gamma function. In recent years the Mittag-Leffler function is gaining more and more popularity in view of the increasing applications of the fractional calculus and is classified as 33E12 in the Mathematics Subject Classification 2000.



also the probability distribution of the jump widths is relevant. The stochastic evolution of the space variable in time is modelled by an integro-differential equation, the *master equation*, which, by containing a time fractional derivative, can be considered as the “fractional” generalization of the Kolmogorov-Feller equation of the compound Poisson process, see <sup>9,13,17</sup>. In particular, in <sup>9</sup> Hilfer and Anton have, without saying it in such words, subordinated a spatially discrete random walk to a “fractional” Poisson renewal process. For more recent results in this respect see <sup>18</sup>, where we have provided the solution of the time fractional master equation in terms of iterated derivatives of a Mittag-Leffler function.

## Acknowledgements

This work has been carried out in the framework of a joint research project for *fractional calculus modelling*, see URL [www.fracalmo.org](http://www.fracalmo.org). R.G. and F.M. appreciate the support of the EU ERASMUS-SOCRATES program for visits to Bologna and Berlin that, besides teaching, were also useful for this research project.

## Appendix A: The Caputo fractional derivative

The *Caputo* fractional derivative provides a fractional generalization of the first derivative through the following rule in the Laplace transform domain,

$$\mathcal{L}\{ {}_t D_*^\beta f(t); s \} = s^\beta \tilde{f}(s) - s^{\beta-1} f(0^+), \quad 0 < \beta \leq 1, \quad s > 0. \quad (\text{A.1})$$

By intending  $f(t)$  be a causal (generalized) function (i.e. with support for  $t \geq 0$ ), it turns out to be, see e.g. <sup>2,8</sup>,

$${}_t D_*^\beta f(t) := \begin{cases} \frac{1}{\Gamma(1-\beta)} \int_0^t \frac{f^{(1)}(\tau)}{(t-\tau)^\beta} d\tau, & 0 < \beta < 1, \\ \frac{d}{dt} f(t), & \beta = 1. \end{cases} \quad (\text{A.2})$$

It can alternatively be written in the form

$$\begin{aligned} {}_t D_*^\beta f(t) &= \frac{1}{\Gamma(1-\beta)} \frac{d}{dt} \int_0^t \frac{f(\tau)}{(t-\tau)^\beta} d\tau - \frac{t^{-\beta}}{\Gamma(1-\beta)} f(0^+) \\ &= \frac{1}{\Gamma(1-\beta)} \frac{d}{dt} \int_0^t \frac{f(\tau) - f(0^+)}{(t-\tau)^\beta} d\tau, \quad 0 < \beta < 1. \end{aligned} \quad (\text{A.3})$$

The Caputo derivative has been indexed with  $*$  in order to distinguish it from the classical Riemann-Liouville derivative  ${}_t D^\beta$ , the first term at the R.H.S. of the first equality in (A.3). As it can be noted from the last equality in (A.3), the Caputo derivative provides a sort of regularization at  $t = 0$  of the Riemann-Liouville derivative; however, it is practically ignored in most mathematical treatises on fractional calculus as <sup>16</sup>. The two notions of fractional derivative can be extended to any order  $\beta > 0$  by introducing the integer  $m$  such that  $m - 1 < \beta \leq m$ ; for details see e.g. <sup>8</sup>. As a conclusive remark, we point out that the *Caputo* derivative of any order  $\beta$  ( $m - 1 < \beta \leq m$ ) satisfies the relevant property of being zero when

applied to a constant, and, in general, to any function  $g_1(t) = \sum_{j=1}^m c_j t^{m-j}$ , like the standard derivative of order  $m$ . On the contrary, the corresponding *Riemann-Liouville* derivative is zero if applied to  $g_2(t) = \sum_{j=1}^m c_j t^{\beta-j}$ , as it turns out from

$${}_t D^\beta t^\gamma = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-\beta)} t^{\gamma-\beta}, \quad \beta > 0, \quad \gamma > -1, \quad t > 0. \quad (A.4)$$

## Appendix B: The Laplace transform of the Mittag-Leffler function

We would like to prove the Laplace transform formula for any  $\beta > 0$

$$\mathcal{L}\{t^{\beta k} E_\beta^{(k)}(\pm at^\beta); s\} = \frac{k! s^{\beta-1}}{(s^\beta \mp a)^{k+1}}, \quad k = 0, 1, 2, \dots, \quad \Re s > |a|^{1/\beta}, \quad (B.1)$$

with  $E_\beta^{(k)}(z) := \frac{d^k}{dz^k} E_\beta(z)$ . In the particular case  $\beta = 1$  the formula reduces to

$$\mathcal{L}\{t^k e^{\pm at}; s\} = \frac{k!}{(s \mp a)^{k+1}}, \quad k = 0, 1, 2, \dots, \quad \Re s > |a|. \quad (B.2)$$

As a matter of fact (B.2) is known to be valid for  $\Re s > \pm \Re a$  and its proof is a consequence of the analyticity property of the Laplace transform,  $\mathcal{L}\{t^k f(t); s\} = (-1)^k \tilde{f}^{(k)}(s)$ , applied to  $f(t) = \exp(\pm at)$ , for which

$$\mathcal{L}\{e^{\pm at}; s\} = \frac{1}{(s \mp a)}, \quad k = 0, 1, 2, \dots, \quad \Re s > \pm \Re a.$$

However, Eq. (B.2) can be deduced for  $\Re s > |a|$  by using the method of power series expansions as shown below. Indeed,

$$\int_0^\infty e^{-u} e^{\pm zu} du = \sum_{k=0}^\infty \frac{(\pm z)^k}{k!} \int_0^\infty e^{-u} u^k du = \sum_{k=0}^\infty (\pm z)^k = \frac{1}{1 \mp z},$$

from which, by differentiating with respect to  $z$ ,

$$\int_0^\infty e^{-u} u^k e^{\pm zu} du = \frac{k!}{(1 \mp z)^{k+1}}, \quad |z| < 1.$$

Now, by introducing the substitutions  $u = st$  and  $z = a/s$  (for our purposes here we agree to take  $s$  real), we get after simple manipulations the identity in (B.2) for  $s > |a|$ . By analytic continuation the validity is extended to complex  $s$  with  $\Re s > |a|$ . The above reasoning can be applied to the integral  $\int_0^\infty e^{-u} E_\beta(\pm zu^\beta) du$  in order to derive the Laplace transform formula (B.1). Indeed,

$$\int_0^\infty e^{-u} E_\beta(\pm zu^\beta) du = \sum_{k=0}^\infty \frac{(\pm z)^k}{\Gamma(\beta k + 1)} \int_0^\infty e^{-u} u^{\beta k} du = \sum_{k=0}^\infty (\pm z)^k = \frac{1}{1 \mp z},$$

from which, by differentiating with respect to  $z$ ,

$$\int_0^\infty e^{-u} u^{\beta k} E_\beta(\pm zu^\beta) du = \frac{k!}{(1 \mp z)^{k+1}}, \quad |z| < 1.$$

Now, by introducing the substitutions  $u = st$  and  $z = a/s^\beta$ , we get after simple manipulations the identity in (B.1) for  $s > |a|^{1/\beta}$ , namely, by analytic continuation, for  $\Re s > |a|^{1/\beta}$ .

## References

1. M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).
2. M. Caputo and F. Mainardi, Linear models of dissipation in anelastic solids, *Riv. Nuovo Cimento* (Ser. II) 1, 161–198 (1971).
3. D.R. Cox, *Renewal Theory*, 2-nd Edn (Methuen, London, 1967).
4. A. Erdélyi, W. Magnus, F. Oberhettinger and F.G. Tricomi, *Higher Transcendental Functions*, Bateman Project, (McGraw-Hill, New York, 1955), Vol 3. [Ch. 18: Miscellaneous Functions, pp. 206–227]
5. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. 2, 2-nd Edn (Wiley, New York, 1971).
6. B.V. Gnedenko and I.N. Kovalenko, *Introduction to Queueing Theory*, (Israel Program for Scientific Translations, Jerusalem, 1968) [Translated from the 1966 Russian edition]
7. R. Gorenflo and E. Abdel-Rehim, From power laws to fractional diffusion, *Vietnam Journal of Mathematics*, to appear.
8. R. Gorenflo and F. Mainardi, Fractional calculus: integral and differential equations of fractional order, in: A. Carpinteri and F. Mainardi (Editors), *Fractals and Fractional Calculus in Continuum Mechanics* (Springer Verlag, Wien, 1997), pp. 223–276. [Reprinted in <http://www.fracalmo.org>]
9. R. Hilfer and L. Anton, Fractional master equations and fractal time random walks, *Phys. Rev. E* 51, R848–R851 (1995).
10. A.Ya. Khintchine, *Mathematical Methods in the Theory of Queueing*, (Charles Griffin, London, 1960). [Translated from the 1955 Russian edition]
11. F. Mainardi, Fractional relaxation-oscillation and fractional diffusion-wave phenomena, *Chaos, Solitons & Fractals* 7, 1461–1477 (1996).
12. F. Mainardi and R. Gorenflo, On Mittag-Leffler type functions in fractional evolution processes, *J. Comput. & Appl. Mathematics* 118, 283–299 (2000).
13. F. Mainardi, M. Raberto, R. Gorenflo and E. Scalas, Fractional calculus and continuous-time finance II: the waiting-time distribution, *Physica A* 287, 468–481 (2000).
14. I. Podlubny, *Fractional Differential Equations* (Academic Press, San Diego, 1999).
15. S.M. Ross, *Introduction to Probability Models*, 6-th Edn (Academic Press, New York, 1997).
16. S.G. Samko, A.A. Kilbas and O.I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach, New York, 1993. Translation from the Russian edition, Nauka i Tekhnika, Minsk, 1987.
17. E. Scalas, R. Gorenflo and F. Mainardi, Fractional calculus and continuous-time finance, *Physica A* 284, 376–384 (2000).
18. E. Scalas, R. Gorenflo and F. Mainardi, Uncoupled continuous-time random walks: solution and limiting behaviour of the master equation, *Physical Review E*, to appear.

# ON THE ACTIVITY OF ABSORBING IRREGULAR INTERFACES

J. S. ANDRADE JR.

*Departamento de Física, Universidade Federal do Ceará,  
60451-970 Fortaleza, Ceará, Brazil*

H. F. DA SILVA

*Departamento de Física, Universidade Federal do Maranhão,  
65080-040 São Luis, Maranhão, Brazil*

E. A. HENRIQUE

*Departamento de Física, Universidade Federal do Ceará,  
60451-970 Fortaleza, Ceará, Brazil*

B. SAPOVAL

*Laboratoire de Physique de la Matière Condensée, CNRS,  
Ecole Polytechnique, 91128 Palaiseau, France  
Centre de Mathématiques et de leurs Applications, CNRS,  
Ecole Normale Supérieure de Cachan, 94235 Cachan, France*

We investigate the activity of  $2d$  absorbing interfaces with irregular geometry under diffusion-limited conditions. First, the theorem of Makarov for Laplacian transport is illustrated with a diffusion cell that includes an absorbing fractal interface. The transition from molecular to Knudsen diffusion on the activity of this cell is then studied through nonequilibrium molecular dynamics simulations. Our results indicate that the extent of the interface that is significantly active is rather sensitive to the governing mechanism of transport. Precisely, the length of the active zone decreases continuously with density from the Knudsen to the molecular diffusion regime. In the limit where molecular diffusion dominates, we find that this length approaches a constant value of the order of the system size, in agreement with theoretical predictions for Laplacian transport in irregular geometries. Finally, we show that all these features can be qualitatively described in terms of a simple random-walk model of the diffusion process.

## 1 Introduction

The phenomenon of Laplacian transport towards irregular surfaces represents a subject of research that has relevance in many fields of technology, including heterogeneous catalysis, heat transfer and electrochemistry. In the case of catalysis, the role of the local surface morphology at the pore level on the activity of the catalyst has been studied with great interest in the past an recent years.<sup>1,2,3</sup> This problem is intimately related with the accessibility of reagent to active sites located along an irregular reactive surface. If the system is diffusion-controlled, *screening* effects may cause a significant reduction on the reactivity of the catalyst surface, as compared to its intrinsic chemical reactivity. In the particular case where the reactivity is extremely high, the surface acts approximately as an idealized source, and the system can be described in terms of Dirichlet (absorbing) boundary conditions. This situation has been thoroughly studied, specially for the case of two-dimensional systems. For instance, an important theorem proposed by Makarov<sup>4</sup> has been used

to describe the properties of the current distribution on irregular electrodes<sup>5</sup> (e.g., fractal electrodes). In terms of catalytic activity, this theorem essentially states that, whatever the shape (perimeter) of the catalyst interface, the size of the region where most of the reaction takes place is of the order of the overall size (or diameter) of the cell under a dilation transformation. Under this framework, several studies have been mainly devoted to the calculation and application of the concept of *active zone* in the Laplacian transport to and across irregular interfaces.<sup>5,6</sup> For example, through the coarse-graining method proposed by Sapoval,<sup>5</sup> it is possible to determine the flux through an arbitrarily irregular surface from its geometry alone, avoiding the solution of the Laplace problem within a complex boundary domain. More recently, it has been shown that this technique provides consistent predictions for the activity of catalyst surfaces.<sup>7</sup>

When dealing with the concept of active zone in Laplacian fields, there is always the implicit assumption that molecular diffusion is the governing mechanism of mass transport. Such an approximation, however, can only be locally valid inside of the void space between the fins or extended protrusions of an irregular surface if the mean free path of the diffusing molecules is sufficiently smaller than the width of these irregularities. As shown in Fig. 1, Knudsen diffusion may become the dominant mechanism of mass transport determining the reactivity of the system if the reagent is a diluted gas for which the collisions among molecules are less frequent than the collisions between the molecules and the catalytic surface.<sup>8,9</sup> The molecular mean free path therefore constitutes a lower cut-off for the validity of the molecular diffusion description.

The aim of the present work is threefold. First, in Section 1, we illustrate the theorem of Makarov by calculating the activity of an absorbing irregular interface subjected to Laplacian transport. Second, in Section 2, we investigate the transition in activity of an irregular absorbing interface when the mechanism of mass transport changes from Knudsen to molecular diffusion. Our approach is to use a nonequilibrium molecular dynamics (NMD) technique in order to simulate a nonuniform and steady-state profile of reagent concentration between two active interfaces with an arbitrarily given roughness. Third, in Section 3, we show that a simple random-walk model of the diffusion-absorption process can provide a consistent description of the behavior observed in the NMD simulations. Finally, some conclusions are drawn in Section 4.

## 2 Laplacian Transport and the Theorem of Makarov

The Laplacian transport phenomenon is illustrated here through direct numerical simulation of diffusion and absorption in a two-dimensional continuum system.<sup>10</sup> For this, we consider the basic cell depicted in Fig. 2 and assume that mass is transported by diffusion from a source at its center line of length  $L$ , towards its fractal interfaces of perimeter  $L_p$ . More precisely, in the bulk of the cell, the transport of mass obeys Fick's law,  $\vec{J}(\vec{r}) \equiv -D\vec{\nabla}C$ , where  $\vec{J}$  represents the mass flux vector field,  $C(\vec{r})$  is the local concentration at position  $\vec{r}$  and  $D$  is the molecular diffusion coefficient. Under steady state conditions, the concentration field satisfies Laplace equation  $\nabla^2 C = 0$ . In addition, a constant unitary concentration is imposed at the

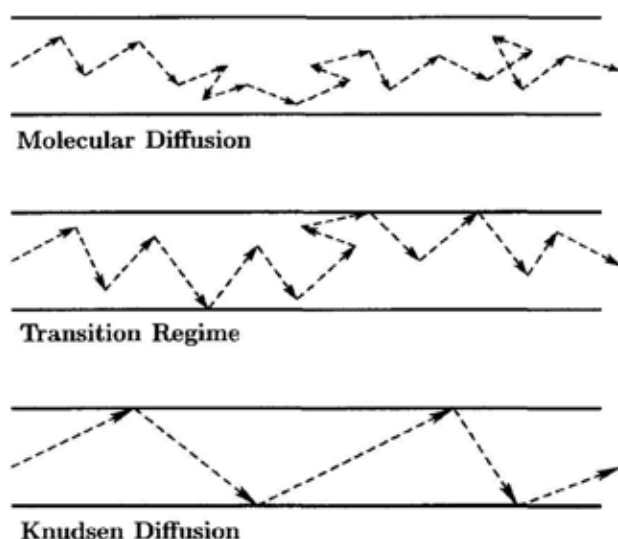


Figure 1. Pictorial representation of different diffusion regimes. On the top, the molecular diffusion regime, where the mean free path is small relative to the pore dimension. In the middle, as the mean free path becomes of the order of the pore diameter, the diffusive species collide with the walls more frequently. Finally, the Knudsen regime is shown on the bottom. In this case, the mean free path is so large compared to the pore diameter, that the collisions between the particles and the wall are more frequent than the collisions among particles.

source line ( $C_0 = 1$ ) and Dirichlet boundary conditions ( $C = 0$ ) are assigned to each elementary unit of the interface. The solution of the Laplacian problem for the concentration field inside the diffusion cell is obtained here through numerical discretization. Due to the symmetry with respect to the source line, only the concentration field in half of the domain needs to be calculated. A structured mesh comprising quadrilateral elements is then generated and the solution is obtained by means of finite-differences.

In Fig. 2 we show the contour plot of the resulting concentration field in logarithmic scale. From the solution, we can compute the local diffusive fluxes  $q_i$  crossing each element  $i$  of the interface. We measure the efficiency of the interface in terms of the active length  $L_a$  defined as<sup>11</sup>

$$L_a \equiv 1 / \sum_{i=1}^{L_p} \phi_i^2 \quad (1 \leq L_a \leq L_p), \quad (1)$$

where the sum is over the total number of interface elements  $L_p$ , and  $\phi_i \equiv q_i / \sum q_j$  is the normalized mass flux at element  $i$  (see Fig. 3). From the definition (1),  $L_a = L_p$  indicates a limiting state of equal partition of fluxes ( $\phi_i = 1/L_p, \forall i$ ) whereas  $L_a = 1$  should correspond to the maximum "localization" of the flux distribution. The calculated active length for the Laplacian cell is found to be

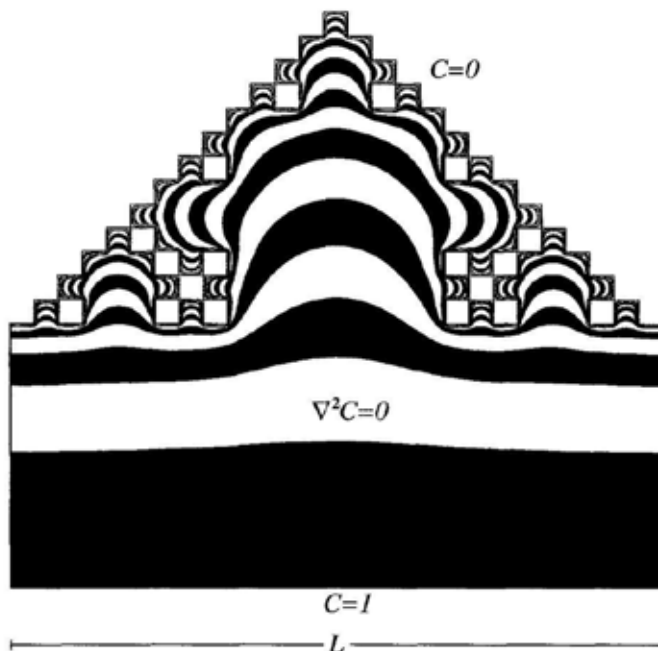


Figure 2. Zebra contour plot of the steady-state concentration field obtained from direct numerical simulation of a Laplacian cell. From the solution of Laplace equation, the logarithm of the concentration is linearly binned in a sequence of black and white stripes (see Ref. [10]). Going from the flat to the fractal interface, the decrease in concentration from one stripe to the next corresponds to a factor of approximately 2.

$L_a = 22.9$ , a value that is much closer to the size of the system  $L = 27$  than to the perimeter of the interface  $L_p = 125$ .

Such a result can be explained in terms of the theorem of Makarov<sup>4</sup>. As already mentioned in the previous section, it describes the properties of Laplacian fields on two-dimensional interfaces of arbitrary shape, subjected to Dirichlet boundary conditions. Precisely, the theorem states that *the information dimension of the harmonic measure on a singly connected interface in  $d = 2$  is exactly equal to 1*. In terms of activity, this means that, regardless the shape of the interface, the total length  $L_a$  of the region where most of the activity takes place should be of the order of the size  $L$  of the cell under a dilation transformation (see Ref. [11] for a detailed discussion of the active zone concept). Therefore, our result for the

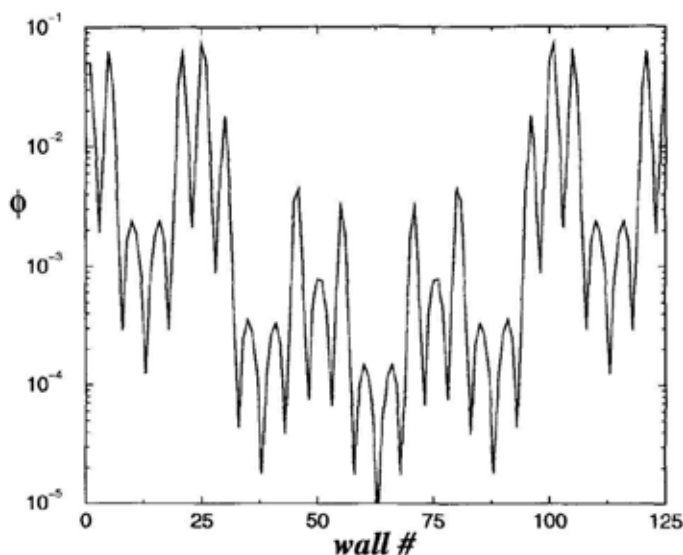


Figure 3. Distribution of the logarithm of the normalized fluxes crossing the wall elements along the absorbing irregular interface of the Laplacian cell.

continuum description of the Laplacian transport is compatible with the prediction of the Makarov theorem,  $L_a \approx L$ .

### 3 The Nonequilibrium Molecular Dynamics Model

In principle, the transition from molecular to Knudsen diffusion can only be predicted in terms of a microdynamical model. For this purpose, we adopt an NMD method that has been originally proposed for the study of self-diffusion in pure fluids.<sup>12</sup> The technique is entirely based on the standard molecular dynamics (MD) at equilibrium, but includes a special scheme to identify and exchange *labeled* and *unlabeled* particles during the simulation.

The MD part of the simulation consists in a two-dimensional cell of size  $L_x l \times L_y l$  containing  $N$  identical particles that interact through the Lennard-Jones potential,  $\Phi(\Delta r_{ij}) = 4\epsilon[(\sigma/\Delta r_{ij})^{12} - (\sigma/\Delta r_{ij})^6]$ , where  $\Delta r_{ij}$  is the distance between particles  $i$  and  $j$ ,  $\epsilon$  is the minimum energy, and  $\sigma$  is the zero of the potential. Periodic boundary conditions are applied in both the  $x$  and  $y$  directions. Distance, energy and time are measured in units of  $\sigma$ ,  $\epsilon$  and  $(m\sigma^2/\epsilon)^{1/2}$ , respectively, and the equations of motion are numerically integrated using the Verlet algorithm.<sup>13</sup>

After thermalization, two identical irregular interfaces of size  $L_y l$  and perimeter  $L_p l$  are symmetrically placed into the system to simulate the roughness geometry of an absorbing material (see Fig. 5). At this point, the non-equilibrium dynamics is put forward through the following scheme: (1) half of the particles in the MD cell are randomly selected to carry a *label*, while the other half are left *unlabeled*; (2) every time a labeled (unlabeled) particle crosses the interface at right (left)



moving in the  $\vec{e}_x$  ( $-\vec{e}_x$ ) direction it becomes unlabeled (labeled), and (3) when reinjected from the right (left) through the periodic boundaries in the  $x$  axis, an unlabeled (labeled) particle becomes labeled (unlabeled). A typical configuration of the system showing the positions of labeled particles is shown in Fig. 4. In Fig. 5 we show the resulting stationary profiles along the  $x$  coordinate of the number fractions  $\theta_l \equiv n_l/(n_l + n_u)$  and  $\theta_u = 1 - \theta_l$ , where  $n_l$  and  $n_u$  are the number of labeled and unlabeled particles, respectively, inside a vertical slice of fixed length in the system. From this point on during the simulation, we keep updating at each time step the number  $n_i$  of particles being “absorbed” by the element  $i$  of the interface in order to compute its local mass flux  $q_i = n_i/\Delta t$ , where  $\Delta t$  is the elapsed time after the steady-state has been established. The active length of the interface is then computed according to the definition (1).

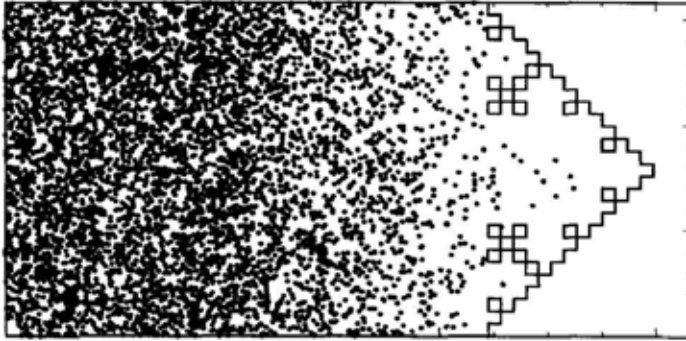


Figure 4. Typical configuration of the NMD system. Only the rough interface at right and the positions of labeled particles are shown.

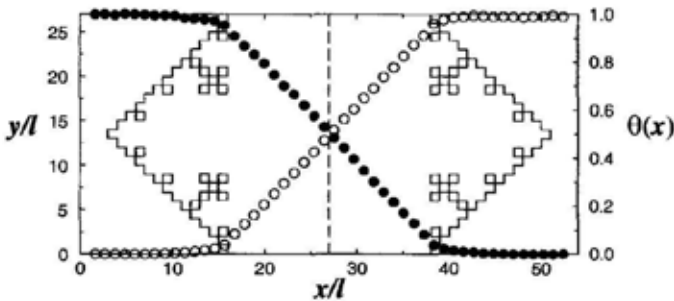


Figure 5. Schematic representation of the NMD diffusion cell. The absorbing interfaces are square Koch trees. Also shown in this figure is the dependence of the local number fraction  $\theta$  of labeled (full circles) and unlabeled (empty circles) particles on the position along the  $x$  direction in the cell. In the case of the random-walk model, particles are released from random  $y$  positions at the dashed line in the center.

Based on this NMD method, we performed simulations for different values of the reduced temperature  $T$ , and reduced densities in the range  $0.025 \leq \rho \leq 0.5$ , corresponding to systems with  $N = 1250$  to 25000 particles. As shown in Fig. 6, the computed active length decreases sharply with  $\rho$  for low density systems at  $T = 1.25$ , up to a point where it remains constant at  $L_a \approx 27$ . The results from simulations performed at a higher temperature,  $T = 3.33$ , show that the behavior of the active length remains nearly the same, at least within the range of densities considered here. The decrease with density of the active length  $L_a$  reflects the transition from Knudsen to molecular diffusion in the distribution of activity at the interface. Because the mean free path of the particles for small  $\rho$  values is larger than the smaller length scale  $l$  of the irregular interface, the activity is highly sensitive to geometrical constraints in the Knudsen regime. At higher densities, the invariant behavior of  $L_a$  is a consequence of molecular diffusion and can be explained in terms of Makarov's theorem.<sup>4</sup> Translating to our diffusion cell, where square Koch trees of third generation are the absorbing interfaces, the theorem of Makarov predicts that the value of  $L_a$  should be close to the size  $L_y = 27$ , in good agreement with the NMD limit obtained for denser systems.

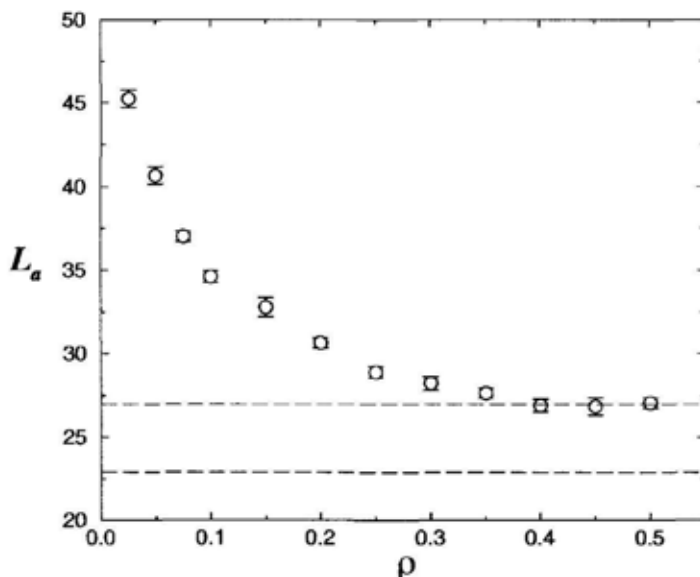


Figure 6. Dependence of the active length  $L_a$  on the reduced density of the NMD cell for a fixed temperature,  $T = 1.25$ . The average values with error bars refer to simulations with 5 different realizations of the NMD process. The horizontal dashed line at the top corresponds to the system size,  $L_y = 27$ , while the one at the bottom indicates the value of the active length obtained from the simulation with the Laplacian cell,  $L_a = 22.9$ .

#### 4 The Random-Walk Model

At this point, we propose a very simple random-walk model that incorporates the basic features of the diffusion-absorption process and is capable to describe, at least semi-quantitatively, the behavior of the active length for different diffusion regimes. Adopting the same geometry of the NMD cell, a particle is released from a random position in the center line. The walker travels through the medium taking steps of random directions, but constant length  $\lambda$ , till it crosses one of the wall elements of the irregular interface and gets absorbed. The flux at this element is then updated and the active length  $L_a$  of the interface recalculated. For a fixed value of the step length  $\lambda$ , the simulation goes on with particles being sequentially released and absorbed, till the active length reaches an average value that is approximately constant. This value is usually obtained with less than  $10^5$  particles launched in the system.

In Fig. 7 we show the dependence on the parameter  $\xi \equiv (\sigma/\lambda)$  of the average  $L_a$  computed for the third generation of the square Koch tree. For a two-dimensional gas,  $\lambda$  can be interpreted as the mean free path, which is inversely proportional to the surface density of the system,  $\lambda \propto 1/\rho$ . Similarly to the NMD simulations, two distinct regimes of activity can be clearly identified and directly related to the different governing mechanisms of mass transport, namely, Knudsen and molecular diffusion. At low values of  $\xi$ , the sharp decrease of  $L_a$  reflects the strong influence on the mass transport process of the irregular geometry of the interface.

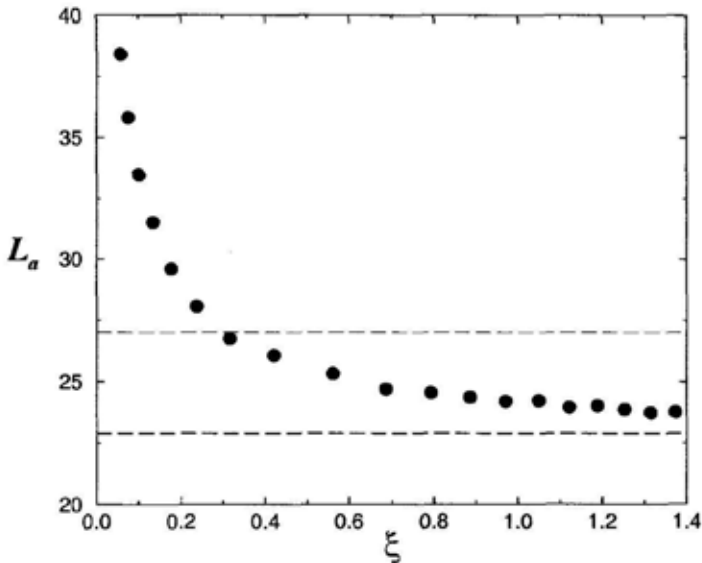


Figure 7. Dependence on the random-walk parameter  $\xi$  of the active length  $L_a$  of the irregular interface for the random-walk model. The horizontal dashed line at the top corresponds to the system size,  $L_y = 27$ , while the one at the bottom gives the active length of the Laplacian cell,  $L_a = 22.9$ .

At sufficiently large values of  $\xi$ , the length  $L_a$  reaches a plateau of minimum activity that is practically coincident with the value of the active length found for the Laplacian cell,  $L_a = 22.9$  (dashed line at the bottom in Fig. 7). Compared to the lower limit of the random-walk model,  $L_a \approx 24$ , the higher value found for the active length with the NMD technique,  $L_a \approx 27$ , can be explained in terms of the structural features and collective behavior of the simulated fluid.

## 5 Summary

In conclusion, we have investigated through molecular dynamics the transition from Knudsen to molecular diffusion transport towards  $2d$  absorbing interfaces with irregular geometry. Our results indicate that the length of the active zone decreases continuously with density from the Knudsen to the molecular diffusion regime. In particular, the active length for absorption of molecular diffusing fluids is found to be very close to that of a purely Laplacian system. Generally speaking, we have shown that the active fraction of an irregular absorbing interface should be sensitive to: (i) its geometrical details; (ii) the governing mechanism of transport, and (iii) the structural aspects of the diffusing fluid. These observations may lead to new guidelines to the problem of diffusion and absorption on arbitrarily irregular interfaces. Furthermore, we have proposed a simple random-walk model that provides substantial insight on the effect of the diffusion mechanism on the interface activity and has the virtue of being computationally cheap. Finally, the approach introduced here is flexible enough to represent specific characteristics of irregular interfaces as well as other types of "absorption" mechanisms (e.g., finite-rate chemical reactions) limited by diffusion transport.

## Acknowledgments

We thank CNPq, CAPES, COFECUB and FUNCAP for support. The Centre de Mathématiques et de leurs Applications and the Laboratoire de Physique de la Matière Condensée are "Unité Mixte de Recherches du Centre National de la Recherche Scientifique" no. 8536 and 7643.

## References

1. R. Gutfraind and M. Sheintuch, *J. Chem. Phys.* **95**, 6100 (1991).
2. M.-O. Coppens, *Catalysis Today* **53**, 225 (1999).
3. M. Sheintuch, *Catalysis Reviews* **43**, 233 (2001).
4. N. G. Makarov, *Proc. London Math. Soc.* **51**, 369 (1985); P. Jones and T. Wolff, *Acta Math.* **161**, 131 (1988).
5. B. Sapoval, *Phys. Rev. Lett.* **73**, 3314 (1994).
6. B. Sapoval, in *Fractals and disordered systems*, edited by A. Bunde and S. Havlin, (Springer-Verlag, Berlin, 1996), 2nd ed., p.232.
7. B. Sapoval, J. S. Andrade Jr., and M. Filoche, *Chem. Eng. Sci.* **56**, 5011 (2001); J. S. Andrade Jr., M. Filoche, and B. Sapoval, *Europhys. Lett.* **55**, 573 (2001).

8. M.-O. Coppens and G. F. Froment, Chem. Eng. Sci. **50**, 1013 (1995); Chem. Eng. Sci. **50**, 1027 (1995).
9. S. B. Santra and B. Sapoval, Phys Rev. E **57**, 6888 (1998).
10. C. J. G. Evertsz and B. B. Mandelbrot B. B., J. Phys. A: Math. Gen. **25** 1781 (1992).
11. B. Sapoval, M. Filoche, K. Karamanos, and R. Brizzi, Eur. Phys. J. B **9**, 739 (1999).
12. J. J. Erpenbeck and W. W. Wood, in *Statistical Mechanics, Part B: Time-dependent Processes*, edited by B. J. Berne, Modern Theoretical Chemistry Vol. 6 (Plenum, New York, 1977); W. Dong and H. Luo, Phys. Rev. E **52**, 801 (1995).
13. M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids*, (Pergamon, Oxford, 1987).

# **FRACTAL DEFORMATION USING DISPLACEMENT VECTORS AND THEIR INCREASING RATES BASED ON EXTENDED UNIT ITERATED SHUFFLE TRANSFORMATION**

TADAHIRO FUJIMOTO AND NORISHIGE CHIBA

*Faculty of Engineering, Iwate University, 4-3-5, Ueda, Morioka, Iwate, 020-8551, Japan*  
*E-mail: {fujimoto,nchiba}@cis.iwate-u.ac.jp*

In this paper, we propose a new fractal deformation technique. An "extended unit Iterated Shuffle Transformation (ext-unit-IST)" is a mapping that changes the order of the places of a code on a code space. When it is applied on a geometric space, it constructs a fractal-like repeated structure, named "local resemblance". In our previously proposed fractal deformation technique, a geometric shape was deformed by applying an ext-unit-IST to displacement vectors (d-vectors) given on the shape. In the new technique proposed in this paper, the ext-unit-IST is applied to the increasing rates of the d-vectors. This allows the d-vectors to change widely without disturbing the shape and improves the deformation quality. Several examples demonstrate the performance of the newly proposed technique.

**Keywords:** computer graphics, geometric model, deformation, IFS, IST.

## 1 Introduction

*Shape deformation* provides not only mathematical interest but also practical usefulness, such as in computer graphics (CG). Common deformation techniques often deform a shape continuously; in CG, such techniques have been proposed a lot.<sup>6</sup> As another type, we have *fractal deformation*. This deforms every subpart of a shape in all scales recursively, and is expected to be useful for creating the shapes of various natural objects. Several fractal deformation techniques have been proposed. Bowman's method<sup>2</sup> deformed an IFS attractor by changing the fixed points and their strength of attraction of its mappings. Burch and Hart<sup>3</sup> proposed a method for maintaining the connectedness of a deformed IFS attractor. Gonzalez<sup>7</sup> proposed a technique for creating a moving fractal tree. Montiel et al.<sup>9</sup> defined and deformed a fractal shape using a recursive functional equation. Sherman and Hart<sup>10</sup> proposed a method for directly manipulating RIFS models. In addition, *fractal interpolation*<sup>1,8</sup> is a topic related to fractal deformation. Zair and Tosan<sup>11</sup> proposed a method that combined IFS and smooth interpolation and achieved fractal deformation.

We proposed a fractal deformation technique, which deformed a shape by giving *displacement vectors* (*d-vectors*) and applying an *extended unit Iterated Shuffle Transformation* (*ext-unit-IST*) to them.<sup>5</sup> This technique handles an original shape and d-vectors for deformation separately, while most other techniques handle the original shape's definition parameters themselves. Thus, our technique enables flexible and intuitive control to achieve desired deformations easily by giving arbitrary d-vectors. Besides, using an ext-unit-IST, 1) the shape is deformed in a fractal-like repeated manner, named "*local resemblance*", 2) it is possible to combine a continuous deformation and a fractal deformation continuously, 3) any shape can be deformed if it is addressed appropriately. However, this technique often disturbs the shape seriously when large d-vectors are given. To avoid this problem, in this paper, we propose a new technique. This technique applies an ext-unit-IST to the

increasing rates of d-vectors instead of d-vectors themselves. As a result, the shape is deformed by changing the expansion rate of area without breaking its topology.

Section 2 gives the definition of ext-unit-IST's, and Sec.3 gives a brief explanation of our fractal deformation technique proposed before; the details of these sections are presented in our previous paper.<sup>5</sup> In Sec.4, we propose a new fractal deformation technique. In Sec.5, we describe the conclusion and future work.

## 2 Extended Unit Iterated Shuffle Transformation

Let  $\Sigma_L$  denote the code space of  $L$  symbols,  $L \geq 1$ , given by

$$\Sigma_L = \{\alpha = \alpha_1 \alpha_2 \alpha_3 \cdots \mid \alpha_j \in \mathbf{Z}_L, j = 1, 2, 3, \dots\}, \quad (1)$$

where  $\mathbf{Z}_L = \{0, 1, \dots, L-1\}$ . We define an ext-unit-IST on  $\Sigma_L$  and obtain some theorems as follows. Their proofs are given in our previous paper.<sup>5</sup>

**Definition 1.** For  $\alpha \in \Sigma_L$ , let  $D_{e-u}^L: \Sigma_L \rightarrow \Sigma_L$  for integers  $k_b \geq 0$  and  $k \geq k_b+1$  be the mapping given by

$$\alpha = \alpha_1 \cdots \alpha_{k_b} \alpha_{k_b+1} \cdots \alpha_{k-1} \alpha_k \alpha_{k+1} \cdots, \quad (2)$$

$$D_{e-u}^L(k_b, k, \alpha) = \alpha_1 \cdots \alpha_{k_b} \alpha_k \alpha_{k_b+1} \cdots \alpha_{k-1} \alpha_{k+1} \cdots. \quad (3)$$

When  $k_b = 0$ , symbols  $\alpha_1, \dots, \alpha_{k_b}$  are omitted from the equations.

**Definition 2.** An ext-unit-IST is defined as the mapping  $D_{e-uni}^L: \Sigma_L \rightarrow \Sigma_L$  for integers  $k_b$  and  $k_e$ ,  $0 \leq k_b \leq k_e$ , given by

$$D_{e-uni}^L(k_b, k_e, \alpha) = \begin{cases} \alpha & \text{if } k_b = k_e, \\ D_{e-u}^L(k_b, k_e, D_{e-uni}^L(k_b, k_e - 1, \alpha)) & \text{if } k_b < k_e. \end{cases} \quad (4)$$

**Theorem 1.** For fixed  $k_b$  and  $k_e$ , the mapping  $D_{e-uni}^L$  is one-to-one and onto.

**Theorem 2.** If  $k_b < k_e$ , then the mapping  $D_{e-uni}^L$  is formulated by

$$\alpha = \alpha_1 \cdots \alpha_{k_b} \alpha_{k_b+1} \alpha_{k_b+2} \cdots \alpha_{k_e-1} \alpha_{k_e} \alpha_{k_e+1} \alpha_{k_e+2} \cdots, \quad (5)$$

$$D_{e-uni}^L(k_b, k_e, \alpha) = \alpha_1 \cdots \alpha_{k_b} \alpha_{k_e} \alpha_{k_e-1} \cdots \alpha_{k_b+2} \alpha_{k_b+1} \alpha_{k_e+1} \alpha_{k_e+2} \cdots. \quad (6)$$

**Theorem 3.** For  $\forall \alpha \in \Sigma_L$ , if  $\alpha' = D_{e-uni}^L(k_b, k_e, \alpha)$  then  $\alpha = D_{e-uni}^L(k_b, k_e, \alpha')$ .

## 3 Fractal Deformation by Applying IST to Displacement Vectors

### 3.1 Addressing Points on Geometric Shape

The fractal deformation of a geometric shape  $S$  is achieved using an ext-unit-IST working on  $S$ . This is realized by giving a code  $\alpha \in \Sigma_L$  to each point  $s \in S$  uniquely as its address using a one-to-one and onto address mapping  $M: S \rightarrow \Sigma_L$ .

$$M(s) = \alpha. \quad (7)$$

Figure 1 (a) shows an address mapping  $M$ ,  $L = 4$ , for a square  $S$ . This illustrates how to give addresses  $\alpha$  on points  $s \in S$ , showing the first and second places of  $\alpha$  given to each region. This addressing proceeds to infinity to give  $\alpha$  infinite places.

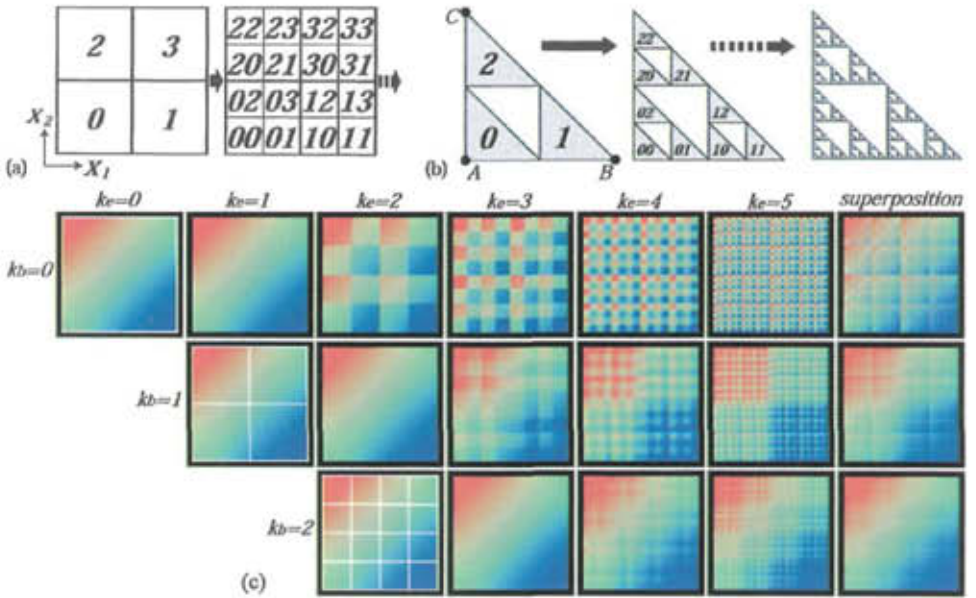


Figure 1. Address mappings and ext-unit-IST's on a geometric shape. (b) is obtained by giving  $a_i = d_i = 0.5$ ,  $b_i = c_i = 0$ ,  $i = 0, 1, 2$ ,  $(e_0, f_0) = (0, 0)$ ,  $(e_1, f_1) = (0.5, 0)$ ,  $(e_2, f_2) = (0, 0.5)$  to Eq.8.

When a shape  $S$  is an IFS (Iterated Function System)<sup>1</sup> attractor, we can utilize the following addressing rule.<sup>1</sup> An IFS consists of a complete metric space  $\mathbf{X}$  and a finite set of contraction mappings  $w_i : \mathbf{X} \rightarrow \mathbf{X}$ ,  $i = 0, \dots, L-1$ . When  $\mathbf{X} = \mathbb{R}^2$ , the mappings  $w_i$  are often given the following affine mapping form:

$$w_i(\mathbf{x}) = w_i \left( \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} e_i \\ f_i \end{bmatrix}, \quad \mathbf{x} \in \mathbb{R}^2. \quad (8)$$

The notation of the IFS is  $\{\mathbf{X}; w_i, i = 0, \dots, L-1\}$ . The set  $S \subset \mathbf{X}$  is referred to as the attractor of the IFS if  $S = \bigcup_{i=0}^{L-1} w_i(S)$ . Then, each point  $s \in S$  is given an address determined in terms of the sequence of mappings  $w_i$  applied. Figure 1 (b) shows the case of a Sierpinski gasket,  $L = 3$ , on  $\mathbf{X} = \mathbb{R}^2$ . The numbers show how to determine addresses  $\alpha$  in the same way as (a).

### 3.2 Ext-unit-IST on Geometric Shape

Using  $D_{e-uni}^L$  in Eq.4 and  $M$  in Eq.7, we define an ext-unit-IST working on a geometric shape  $S$  as the mapping  $F_{e-uni}^L : S \rightarrow S$  for integers  $k_b$  and  $k_e$ .

$$F_{e-uni}^L(k_b, k_e, s) = M^{-1}(D_{e-uni}^L(k_b, k_e, M(s))). \quad (9)$$

Each point  $s \in S$  can be given various attributes. So, let  $F_a : S \rightarrow Atr$  be an attribute function that gives an attribute value  $a \in Atr$  to a point  $s$ .

$$F_a(s) = a. \quad (10)$$



For example, a *coordinate function*  $F_{\mathbf{x}} : S \rightarrow \mathbf{X}$  gives a coordinate  $\mathbf{x} \in \mathbf{X}$  to a point  $s \in S$ , where  $\mathbf{X}$  is the space on which  $S$  is defined.

$$F_{\mathbf{x}}(s) = \mathbf{x}. \quad (11)$$

Attribute values  $a$  given to points  $s$  by the attribute function  $F_a$  are transformed on the shape  $S$  by the application of the ext-unit-IST. Using Equations 9 and 10, the resultant function  $F_{e-\text{uni},a}^L : S \rightarrow Atr$  is expressed as follows.

$$F_{e-\text{uni},a}^L(k_b, k_e, s) = F_a(F_{e-\text{uni}}^{L-1}(k_b, k_e, s)). \quad (12)$$

Moreover, Eq.12 can be *superposed* as follows.

$$F_{S-e-\text{uni},a}^L(k_b, k_e, s) = \sum_{k=k_b}^{k_e} W_a(k, k_b, k_e, s) F_{e-\text{uni},a}^L(k_b, k, s). \quad (13)$$

The *weight function*  $W_a$ ,  $k_b \leq k \leq k_e$ , is defined using a *ratio function*  $\delta_a(s) \geq 0$ .

$$W_a(k, k_b, k_e, s) = \begin{cases} \{1 - \delta_a(s)\} \{\delta_a(s)\}^{k-k_b} / [1 - \{\delta_a(s)\}^{k_e-k_b+1}] & \text{if } \delta_a(s) \neq 1, \\ 1/(k_e - k_b + 1) & \text{if } \delta_a(s) = 1. \end{cases} \quad (14)$$

Equation 14 satisfies  $\sum_{k=k_b}^{k_e} W_a(k, k_b, k_e, s) \equiv 1$ . The weight function and ratio function are given  $s$  as an argument so as to be variable for different points  $s \in S$ .

An ext-unit-IST constructs the structure of “local resemblance in space/scale directions”<sup>4,5</sup> on a geometric space. This structure is a combination of “locality in space directions” in Euclidean geometry and “self-similarity in scale directions” in fractal geometry, where *space directions* are the directions along which a point of view changes, and *scale directions* are those along which a field of view changes.

Figure 1 (c) shows how ext-unit-IST’s work on a geometric shape  $S$ , a square, in local resemblance manner when different pairs of  $k_b$  and  $k_e$  are given. The shape  $S$  is given the address mapping of Fig.1 (a). A point  $s \in S$  is given a color value  $\mathbf{c} = (R, G, B) \in \mathbf{C}$  by a *color function*  $F_{\mathbf{c}} : S \rightarrow \mathbf{C}$ . The image in the position  $(k_b, k_e)$  is colored by  $F_{e-\text{uni},\mathbf{c}}^L$  using the values  $k_b$  and  $k_e$  (cf. Eq.12). Each image in the column “superposition” is colored by  $F_{S-e-\text{uni},\mathbf{c}}^L$  using  $k_b$  of its row,  $k_e = 5$ , and  $\delta_{\mathbf{c}}(s) = 1$  for all  $s$  (cf. Eq.13). Each image has a different fractal-like repeated structure. The value  $k_e$  controls the level of detail of the repeated structure. For a fixed  $k_b$ , as  $k_e$  increases, the structure is constructed into the scale direction by scattering colors over the square; even though  $k_e$  increases, the color proportion of  $R, G, B$  depends on the position in the space direction. This shows the “local resemblance in space/scale directions” constructed by the ext-unit-IST’s. The value  $k_b$  restricts the extent within which a point  $s$  can be transformed. For each  $k_b$ , the leftmost image shows that a point  $s$  can be transformed only within the region that is enclosed by a white square and includes the point. Comparing the cases of  $k_b = 0, 1, 2$  shows that, as  $k_b$  becomes greater, the scattering of colors is localized within smaller regions. This property realizes a continuous transition between a continuous deformation and a fractal deformation (cf. Sec.4.2).

### 3.3 Fractal Deformation Using Ext-unit-IST to D-vectors

The fractal deformation proposed in our previous paper<sup>5</sup> is achieved by giving displacement vectors (d-vectors) to the points on a geometric shape to deform and

applying an ext-unit-IST to the d-vectors. Let  $F_v : S \rightarrow \mathbf{X}$  be a *displacement vector function* (d-vector function) to give a d-vector  $\mathbf{v} \in \mathbf{X}$  to each point  $s$  on a shape  $S$ .

$$F_v(s) = \mathbf{v}. \quad (15)$$

When  $F_v$  is directly added to the coordinate function  $F_x$  of  $S$ , the coordinate of a point  $s \in S$  is obtained by the following function  $P_{x,v} : S \rightarrow \mathbf{X}$ .

$$P_{x,v}(s) = F_x(s) + F_v(s). \quad (16)$$

We consider  $F_v$  such that  $\mathbf{v}$  varies continuously on the space  $\mathbf{X}$ , which is the range of  $F_x$ . In this case, by Eq.16,  $S$  is deformed continuously. We refer to this deformation as *continuous deformation*. On the other hand, the fractal deformation we proposed<sup>5</sup> is defined using Eq.9 as the following function  $P_{e-\text{uni}, x, v}^L : S \rightarrow \mathbf{X}$ .

$$\begin{aligned} P_{e-\text{uni}, x, v}^L(k_b, k_e, s) &= F_x(s) + F_v(F_{e-\text{uni}}^L{}^{-1}(k_b, k_e, s)) \\ &= F_x(s) + F_{e-\text{uni}, v}^L(k_b, k_e, s). \end{aligned} \quad (17)$$

In Eq.17,  $F_{e-\text{uni}}^L$  rearranges the d-vectors given by  $F_v$  on  $S$  in local resemblance manner (cf. Sec.3.2). Using Eq.13, the superposed case is obtained as follows.

$$\begin{aligned} P_{S-e-\text{uni}, x, v}^L(k_b, k_e, s) &= F_x(s) + \sum_{k=k_b}^{k_e} W_v(k, k_b, k_e, s) F_{e-\text{uni}, v}^L(k_b, k, s) \\ &= F_x(s) + F_{S-e-\text{uni}, v}^L(k_b, k_e, s). \end{aligned} \quad (18)$$

The weight function  $W_v$  for  $\mathbf{v}$  is obtained by a ratio function  $\delta_v$  (cf. Eq.14).

The transformations above are easily extended for non-integers  $k_b^*$  and  $k_e^*$  by interpolating the four results obtained for  $([k_b^*], [k_e^*])$ ,  $([k_b^*], [k_e^*] + 1)$ ,  $([k_b^*] + 1, [k_e^*])$ , and  $([k_b^*] + 1, [k_e^*] + 1)$ , where  $[x]$  means the integral part of a non-integer  $x$ .

#### 4 Fractal Deformation by Applying IST to Increasing Rates of Displacement Vectors

In this section, we propose a new fractal deformation technique. When using the previous technique described in Sec.3, the movements of points  $s$  on a geometric shape  $S$  can be overlapped (cf. Fig.2 (c-1, ..., 4), Fig.3 (c-1, ..., 4)), because each point  $s$  is moved only by the d-vector given to the point independent of the movements of other points. Thus, when large d-vectors are given, the shape is seriously disturbed. To avoid this problem, the new technique proposed in this paper transforms the increasing rates of d-vectors, which are obtained by considering the neighboring d-vectors, instead of d-vectors themselves. This results in transforming the expansion rate of area on the shape without breaking its original topology.

In the following, we treat the case when  $\mathbf{X} = \mathbf{R}^2$ ,  $\mathbf{x} = (x, y) \in \mathbf{R}^2$ , and  $\mathbf{v} = (v_x, v_y) \in \mathbf{R}^2$ . For convenience, in the following description, we use  $a$  instead of  $v_x$  and  $v_y$ . The  $a$ 's that appear in the following can be replaced with  $v_x$ 's or  $v_y$ 's.

##### 4.1 Fractal Deformation Using Ext-unit-IST to Increasing Rates of D-vectors

Let  $F_{\partial a / \partial x}(s)$  and  $F_{\partial a / \partial y}(s)$  denote the functions that give the increasing rates of an attribute value  $a$ , which is given by  $F_a(s)$  of Eq.10, in the  $x$  and  $y$  directions.

Then, applying an ext-unit-IST to these functions makes the following (cf. Eq.12).

$$F_{e-\text{uni}, \partial a / \partial x}^L(k_b, k_e, s) = F_{\partial a / \partial x}(F_{e-\text{uni}}^{L-1}(k_b, k_e, s)), \quad (19)$$

$$F_{e-\text{uni}, \partial a / \partial y}^L(k_b, k_e, s) = F_{\partial a / \partial y}(F_{e-\text{uni}}^{L-1}(k_b, k_e, s)). \quad (20)$$

The superposed cases are the following.

$$F_{S-e-\text{uni}, \partial a / \partial x}^L(k_b, k_e, s) = \sum_{k=k_b}^{k_e} W_{\partial a / \partial x}(k, k_b, k_e, s) F_{e-\text{uni}, \partial a / \partial x}^L(k_b, k, s), \quad (21)$$

$$F_{S-e-\text{uni}, \partial a / \partial y}^L(k_b, k_e, s) = \sum_{k=k_b}^{k_e} W_{\partial a / \partial y}(k, k_b, k_e, s) F_{e-\text{uni}, \partial a / \partial y}^L(k_b, k, s). \quad (22)$$

Our goal is to obtain the function  $F_{e-\text{uni}, a}^{L, \text{rat}}(k_b, k_e, s)$  whose increasing rates are equal to Equations 19 and 20, or  $F_{S-e-\text{uni}, a}^{L, \text{rat}}(k_b, k_e, s)$  to Equations 21 and 22. We adopt a numerical way. We take  $N$  calculation points (*c-points*)  $s_i$ ,  $i = 1, \dots, N$ , on the shape  $S$  and obtain the increasing rates  $Fax_i$  and  $Fay_i$  on these *c-points*.

$$Fax_i = F_{e-\text{uni}, \partial a / \partial x}^L(k_b, k_e, s_i) \quad \text{or} \quad F_{S-e-\text{uni}, \partial a / \partial x}^L(k_b, k_e, s_i), \quad (23)$$

$$Fay_i = F_{e-\text{uni}, \partial a / \partial y}^L(k_b, k_e, s_i) \quad \text{or} \quad F_{S-e-\text{uni}, \partial a / \partial y}^L(k_b, k_e, s_i). \quad (24)$$

It is generally difficult to obtain attribute values  $a_i$  on the *c-points*  $s_i$  using common numerical integration methods, because the shape  $S$  does not always have a continuous domain and is hard to determine a path of integration on, such as usual fractal shapes. Therefore, we try to obtain an approximate solution; we find a set of  $a_i$  on all the *c-points*  $s_i$  so that the increasing rates on each  $s_i$  calculated from the set of  $a_i$  are close to  $Fax_i$  and  $Fay_i$  as much as possible (\*).

We approximately represent the increasing rates of  $a$  on each *c-point*  $s_i$  using a set of  $a_i$  as follows. We consider that each  $s_i$  has a coordinate  $\mathbf{p}_i = (x_i, y_i, a_i)$  on the three-dimensional space  $x$ - $y$ - $a$ . Using  $(x_i, y_i)$ , a two-dimensional Delaunay triangulation can be constructed on the  $x$ - $y$  plane. This makes a triangulated polygonal surface on the space  $x$ - $y$ - $a$ . Then, we define the normal vector  $\mathbf{n}_i$  of  $s_i$  on the surface. First, let  $s_{i_k}$ ,  $k = 1, \dots, N_i$ , denote the  $N_i$  *c-points* connected to  $s_i$  on the Delaunay triangulation, where the *c-points* from  $s_{i_1}$  to  $s_{i_{N_i}}$  are placed counterclockwise around  $s_i$  on the  $x$ - $y$  plane. The coordinate of  $s_{i_k}$  is  $\mathbf{p}_{i_k} = (x_{i_k}, y_{i_k}, a_{i_k})$ . Then, using the relative vector  $\mathbf{r}_{i_k} = \mathbf{p}_{i_k} - \mathbf{p}_i = (rx_{i_k}, ry_{i_k}, ra_{i_k})$ , the normal vector  $\mathbf{n}_{i_k}$  of each triangle polygon around  $s_i$  is represented by the following outer product.

$$\mathbf{n}_{i_k} = \mathbf{r}_{i_k} \times \mathbf{r}_{i_{[k+1]}} = (nx_{i_k}, ny_{i_k}, na_{i_k}), \quad k = 1, \dots, M_i. \quad (25)$$

If  $s_i$  is an inner point on the polygonal surface, then  $s_i$  has  $N_i$  triangles and  $M_i = N_i$ . The notation  $[k+1]$  means that if  $k \neq N_i$  then  $[k+1] = k+1$  and if  $k = N_i$  then  $[k+1] = 1$ . If  $s_i$  is a boundary point, then  $s_i$  has  $N_i - 1$  triangles and  $M_i = N_i - 1$ . Using Eq.25, we define the normal vector  $\mathbf{n}_i$  of  $s_i$  as the sum of  $\mathbf{n}_{i_k}$ .<sup>a</sup>

$$\mathbf{n}_i = \sum_{k=1}^{M_i} \mathbf{n}_{i_k} = (nx_i, ny_i, na_i). \quad (26)$$

<sup>a</sup>The normal vectors  $\mathbf{n}_{i_k}$  in Eq.25 are not normalized. This means that not only the direction but also the length of each  $\mathbf{n}_{i_k}$  affects  $\mathbf{n}_i$ . We need to investigate what occurs by this definition.

Each element of  $\mathbf{n}_i$  is represented as follows. If the c-point  $s_i$  is an inner point, then

$$nx_i = \sum_{k=1}^{N_i} (y_{i[k-1]} - y_{i[k+1]})a_{i_k}, \quad (27)$$

$$ny_i = \sum_{k=1}^{N_i} (x_{i[k+1]} - x_{i[k-1]})a_{i_k}, \quad (28)$$

$$na_i = \sum_{k=1}^{N_i} (x_{i_k}y_{i[k+1]} - y_{i_k}x_{i[k+1]}), \quad (29)$$

where if  $k \neq 1$  then  $[k-1] = k-1$  and if  $k = 1$  then  $[k-1] = N_i$ . If the c-point  $s_i$  is a boundary point, then

$$\begin{aligned} nx_i = & \sum_{k=2}^{N_i-1} (y_{i_{k-1}} - y_{i_{k+1}})a_{i_k} \\ & + (y_{i_{N_i}} - y_{i_1})a_{i_1} + (y_{i_2} - y_{i_1})a_{i_1} + (y_{i_{N_i-1}} - y_{i_{N_i}})a_{i_{N_i}}, \end{aligned} \quad (30)$$

$$\begin{aligned} ny_i = & \sum_{k=2}^{N_i-1} (x_{i_{k+1}} - x_{i_{k-1}})a_{i_k} \\ & + (x_{i_1} - x_{i_{N_i}})a_{i_1} + (x_{i_2} - x_{i_1})a_{i_1} + (x_{i_{N_i-1}} - x_{i_{N_i}})a_{i_{N_i}}, \end{aligned} \quad (31)$$

$$\begin{aligned} na_i = & \sum_{k=1}^{N_i-1} (x_{i_k}y_{i_{k+1}} - y_{i_k}x_{i_{k+1}}) \\ & + (x_{i_{N_i}}y_{i_1} - y_{i_{N_i}}x_{i_1}). \end{aligned} \quad (32)$$

Considering the relation between normal and tangential vectors on the space  $x$ - $y$ - $a$  gives the increasing rates  $Dax_i$  and  $Day_i$  of  $a$  on  $s_i$  in the  $x$  and  $y$  directions.

$$Dax_i = -nx_i/na_i, \quad Day_i = -ny_i/na_i. \quad (33)$$

To achieve our goal ((\*) above), we consider the following error function.

$$E_a = \sum_{i=1}^N \{(Fax_i - Dax_i)^2 + (Fay_i - Day_i)^2\}. \quad (34)$$

Determining the values of  $a_i$ ,  $i = 1, \dots, N$ , so as to minimize  $E_a$  achieves our goal. We use the method of least squares; we solve the system of linear equations  $\partial E_a / \partial a_j = 0$ ,  $j = 1, \dots, N$ , with  $N$  unknowns  $a_i$ ,  $i = 1, \dots, N$ . These equations are transformed to the following form using the fact that  $nx_i$  and  $ny_i$  include  $a_i$  as variables while  $na_i$  do not (cf. Equations 27 to 32).

$$\sum_{i=1}^N \{nx_i \cdot \frac{\partial nx_i}{\partial a_j} + ny_i \cdot \frac{\partial ny_i}{\partial a_j}\} = - \sum_{i=1}^N na_i \{Fax_i \cdot \frac{\partial nx_i}{\partial a_j} + Fay_i \cdot \frac{\partial ny_i}{\partial a_j}\}. \quad (35)$$

In order to complete Eq.35, the values of  $\partial nx_i / \partial a_j$  and  $\partial ny_i / \partial a_j$  have to be determined. Using Equations 27, 28, 30, and 31, they are obtained as follows.

- (1) The case when  $i = j$  : If the c-point  $s_i$  is an inner point, then

$$\partial nx_i / \partial a_j = 0, \quad \partial ny_i / \partial a_j = 0. \quad (36)$$

If the c-point  $s_i$  is a boundary point, then

$$\partial nx_i / \partial a_j = y_{i_{N_i}} - y_{i_1}, \quad \partial ny_i / \partial a_j = x_{i_1} - x_{i_{N_i}}. \quad (37)$$

- (2) The case when the c-point  $s_i$  is connected with the c-point  $s_j$  :

If the c-point  $s_i$  is an inner point, then

$$\partial nx_i / \partial a_j = y_{i_{[m-1]}} - y_{i_{[m+1]}}, \quad \partial ny_i / \partial a_j = x_{i_{[m+1]}} - x_{i_{[m-1]}}, \quad \text{where } j = i_m. \quad (38)$$

If the c-point  $s_i$  is a boundary point, then

$$(i) \text{ If } j = i_1, \text{ then } \partial n x_i / \partial a_j = y_i - y_{i_2}, \partial n y_i / \partial a_j = x_{i_2} - x_i. \quad (39)$$

$$(ii) \text{ If } j = i_{N_i}, \text{ then } \partial n x_i / \partial a_j = y_{i_{N_i-1}} - y_i, \partial n y_i / \partial a_j = x_i - x_{i_{N_i-1}}. \quad (40)$$

(iii) If  $j = i_m \neq i_1, i_{N_i}$ , then

$$\partial n x_i / \partial a_j = y_{i_{m-1}} - y_{i_{m+1}}, \partial n y_i / \partial a_j = x_{i_{m+1}} - x_{i_{m-1}}. \quad (41)$$

(3) The case when the c-point  $s_i$  is not connected with the c-point  $s_j$  :

$$\partial n x_i / \partial a_j = 0, \partial n y_i / \partial a_j = 0. \quad (42)$$

Actually, the solution of Eq.35 cannot be determined because the  $N$  equations are not independent by the fact each equation is based on relative vectors  $\mathbf{r}_{ik}$ . So, we solve the equations as a system of linear equations with  $N - 1$  unknowns  $a_i$ ,  $i = 1, \dots, N - 1$ , by giving  $a_N = 0$ . Then, we choose  $N_f$  c-points  $s_{ft(l)}$ ,  $l = 1, \dots, N_f$ ,  $1 \leq ft(l) \leq N$ , as *fitting points* and give attribute values  $a'_{ft(l)}$  as constraints. The attribute values  $a'_i$  of other c-points  $s_i$  are determined by

$$a'_i = a_i + \sum_{l=1}^{N_f} w_{ft}(s_i, s_{ft(l)})(a'_{ft(l)} - a_{ft(l)}), \quad i \neq ft(l), \quad l = 1, \dots, N_f, \quad (43)$$

where  $a_i$  and  $a_{ft(l)}$  are the solutions above,  $w_{ft}$  gives a weight value inversely proportional to the distance between  $s_i$  and  $s_{ft(l)}$ , and  $\sum_{l=1}^{N_f} w_{ft}(s_i, s_{ft(l)}) \equiv 1$ . We finally obtain the objective function  $F_{e-\text{uni}, a}^{L, \text{rat}}$  or  $F_{S-e-\text{uni}, a}^{L, \text{rat}}$  by interpolating the values  $a'_i$ ,  $i = 1, \dots, N$ , on the Delaunay triangulation of  $(x_i, y_i)$ . We actually obtain  $F_{e-\text{uni}, \mathbf{v}}^{L, \text{rat}} = (F_{e-\text{uni}, v_x}^{L, \text{rat}}, F_{e-\text{uni}, v_y}^{L, \text{rat}})$  or  $F_{S-e-\text{uni}, \mathbf{v}}^{L, \text{rat}} = (F_{S-e-\text{uni}, v_x}^{L, \text{rat}}, F_{S-e-\text{uni}, v_y}^{L, \text{rat}})$ .

## 4.2 Examples

Figures 2 and 3 are examples produced by the proposed technique.

▷ The animations of these examples are shown at the following web site.

<http://www-cg.cis.iwate-u.ac.jp/~fujimoto/frac04/fujimoto-frac04-fig.html>

They help readers understand the following explanation.

In Fig.2, (a) is an IFS attractor of  $L = 4$  named “fractal square (f-square)”. This is treated as  $F_{\mathbf{x}}$  and given the same addressing rule as Fig.1 (a). (b) is a continuous deformation by Eq.16. (c-1, ..., 4) are fractal deformations by Eq.18 of the previous technique. (d-1, ..., 4), (e), (f) are fractal deformations by the new technique proposed in Sec.4.1 (superposed cases). These fractal deformations are given different pairs of  $k_b$  and  $\delta_{\mathbf{v}}$ , the same  $k_e = 6$ , and the same d-vector function  $F_{\mathbf{v}}$  as (b). Technically,  $F_{\mathbf{v}}$  was defined by giving d-vectors to  $4 \times 4$  control points over the f-square and applying Bézier interpolation, which is represented by the white grid in each image. For (d-1), ..., (f),  $N = 4^6 = 4096$  c-points were used (these were also used for (b), ..., (c-4)). The small white squares in the images indicate  $L = 4$  fixed points used as fitting points  $s_{ft(l)}$ ,  $l = 1, \dots, 4$ , which are given constraints  $F_{S-e-\text{uni}, \mathbf{v}}^L(k_b, k_e, s_{ft(l)})$ ; the same applies to the cases in Fig.3. For displaying the shapes, plenty of points were produced by interpolating the c-points.

In (c-1, ..., 4),  $F_{\mathbf{v}}$  gives too large d-vectors to keep the deformations stable; the movements of points overlap one another, and the deformed shapes are disturbed. On the other hand, in (d-1), ..., (f), transforming the increasing rates of d-vectors

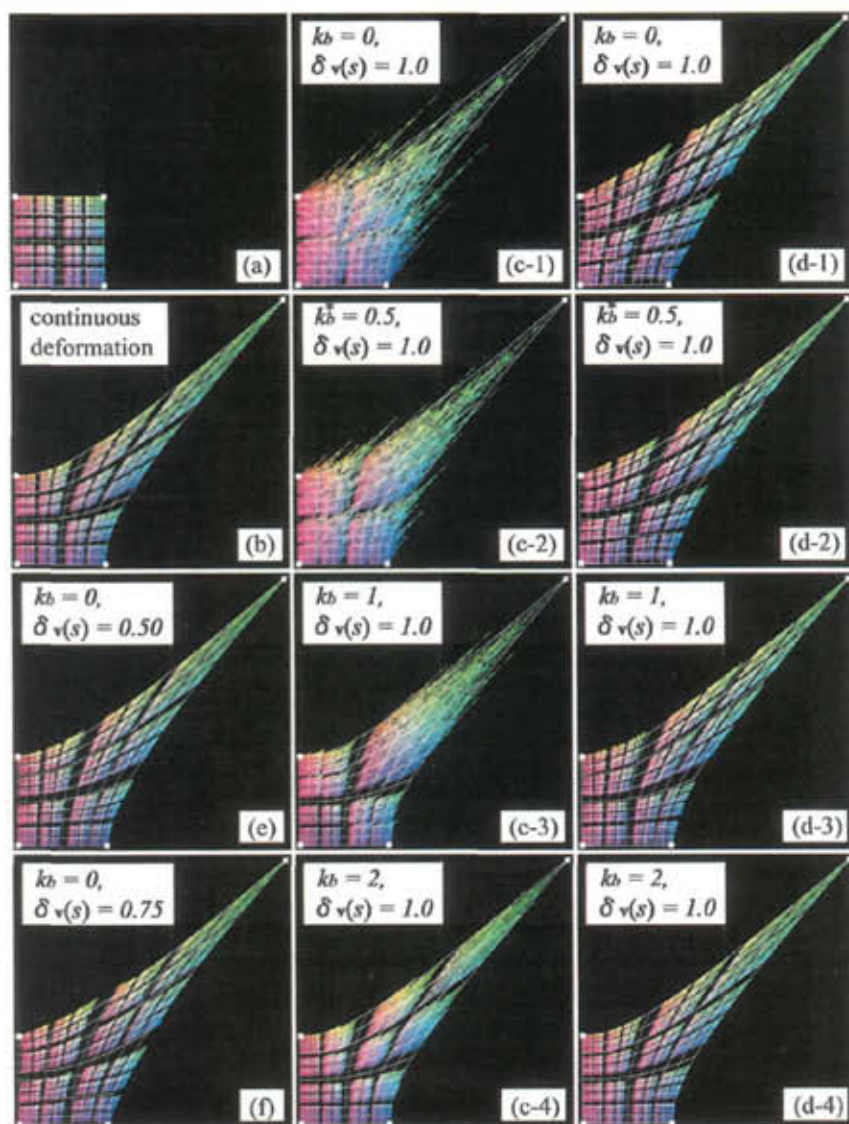


Figure 2. Fractal deformations (1). (a) is defined by  $a_i = d_i = 0.45$ ,  $b_i = c_i = 0$ ,  $i = 0, \dots, 3$ ,  $(e_0, f_0) = (0, 0)$ ,  $(e_1, f_1) = (0.55, 0)$ ,  $(e_2, f_2) = (0, 0.55)$ ,  $(e_3, f_3) = (0.55, 0.55)$  in Eq.8.

results in replacing the expansion rate of area in the vicinity of a point with that of another point; on the deformed shapes, fractal-like repeated changes are given to the expansion rate of area without breaking their original topology.

The continuous deformation (b) deforms the whole f-square continuously. The draw of the upper right (UR-) corner is propagated around continuously; the neighbourhood of the corner is deformed greater, and far regions are less deformed; the



region near the lower left (LL-) corner is hardly deformed. On the other hand, the fractal deformations (d-1), ..., (f) deform the f-square in the way that each subpart is deformed recursively. The draw of the UR-corner is scattered all over the f-square in fractal-like repeated manner; every sub-f-square in all scales and all regions, even near the LL-corner, is deformed, although the sub-f-squares near the UR-corner are still deformed greater than those in far regions. This deformation manner indicates the property of local resemblance.

In (d-1, ..., 4), as  $k_b$  increases from 0 to 2, the sub-f-squares near the UR-corner get to deform greater while those in other regions get to deform smaller. This means that the scattering of the draw of the UR-corner is localized into each original position on the f-square (cf. Fig.1 (c)). When  $k_b$  reaches  $k_e$ , this localization converges; the fractal deformations by  $F_{S-e-uni, v}^{L, rat}$  become the continuous deformation by  $F_v$ , and we obtain the same result as (b). This is easily understood by Eq.4; if  $k_b = k_e$ , an ext-unit-IST gives no change to  $\alpha$ . Thus, an ext-unit-IST enables a continuous transition between a continuous deformation and a fractal deformation by changing  $k_b$  continuously (cf. Sec.3.2). By using non-integer cases, such as (d-2), the continuous transition is perfectly realized (cf. Sec.3.3).

In (d-1, ..., 4),  $\delta_v(s)$  is set to 1.0 <sup>b</sup> for all  $s$  to make the deformation effects clearly understandable. In (e) and (f),  $\delta_v(s)$  is set to 0.5 and 0.75. As  $\delta_v(s)$  becomes greater, the movements of points in smaller scales stand out. This effect is changeable according to positions by making  $\delta_v(s)$  vary over  $s$ .

The images from (a-0) to (e) in Fig.3 show Sierpinski gaskets,  $L = 3$ . (b) is a continuous deformation by Eq.16. (c-1, ..., 4) are fractal deformations by Eq.18, while (d-1, ..., 4) and (e) are fractal deformations by the new technique. They are given the same  $F_v$  as (b) and the same  $k_e = 7$ . (a-0) shows  $N = 3^7 = 2187$  c-points. (a-1, 3, 4) show the regions within which points can move in (c, d-1, 3, 4) in the same way as Fig.1 (c). Compared with (c-1, ..., 4), the shapes in (d-1, ..., 4) and (e) are deformed without being broken. Besides, every triangle in all scales and regions, even near the left side edge, on the gasket is deformed in local resemblance manner in (d-1, ..., 4) and (e), while the whole gasket is continuously deformed and triangles near the left side edge are hardly deformed in (b). Because a Sierpinski gasket has large holes, the increasing rates by Eq.33 based on the Delaunay triangulation of c-points lack correctness at some points, particularly near large holes. This made the disturbances near large holes in (d-1, ..., 4) and (e), although these disturbances were reduced by averaging the increasing rates in the vicinity of each point.

The images from (f-1) to (k-2) in Fig.3 show Twin-dragons,  $L = 2$ . (f-1, 2) show the addressing rule. For the original shape (g), (h-1, 2) are continuous deformations by Eq.16. (i, j, k-1, 2) are fractal deformations by the new technique. (h, i, j, k-1) are given the same  $F_v$ , and so are (h, i, j, k-2).  $N = 2^{12} = 4096$  c-points were used. In these cases,  $F_{e-uni, \partial\alpha/\partial x}^L(k_b, k, s)$  in Eq.21 and  $F_{e-uni, \partial\alpha/\partial y}^L(k_b, k, s)$  in Eq.22 are rotated by the rotation factors of the IFS mappings from  $w_{\alpha_{k_b+1}}$  to  $w_{\alpha_k}$ , where  $\alpha = M(s)$  by Eq.7. This provides spiral deformation effects. In (i-1), the movement of the right part of the dragon, shown in (h-1), is scattered everywhere, resulting in deforming spirally every tip in all scales and regions, even on the opposite side.

<sup>b</sup>This means that the weights in all scales are the same, although a fractal shape usually reduces its weights into smaller scales.

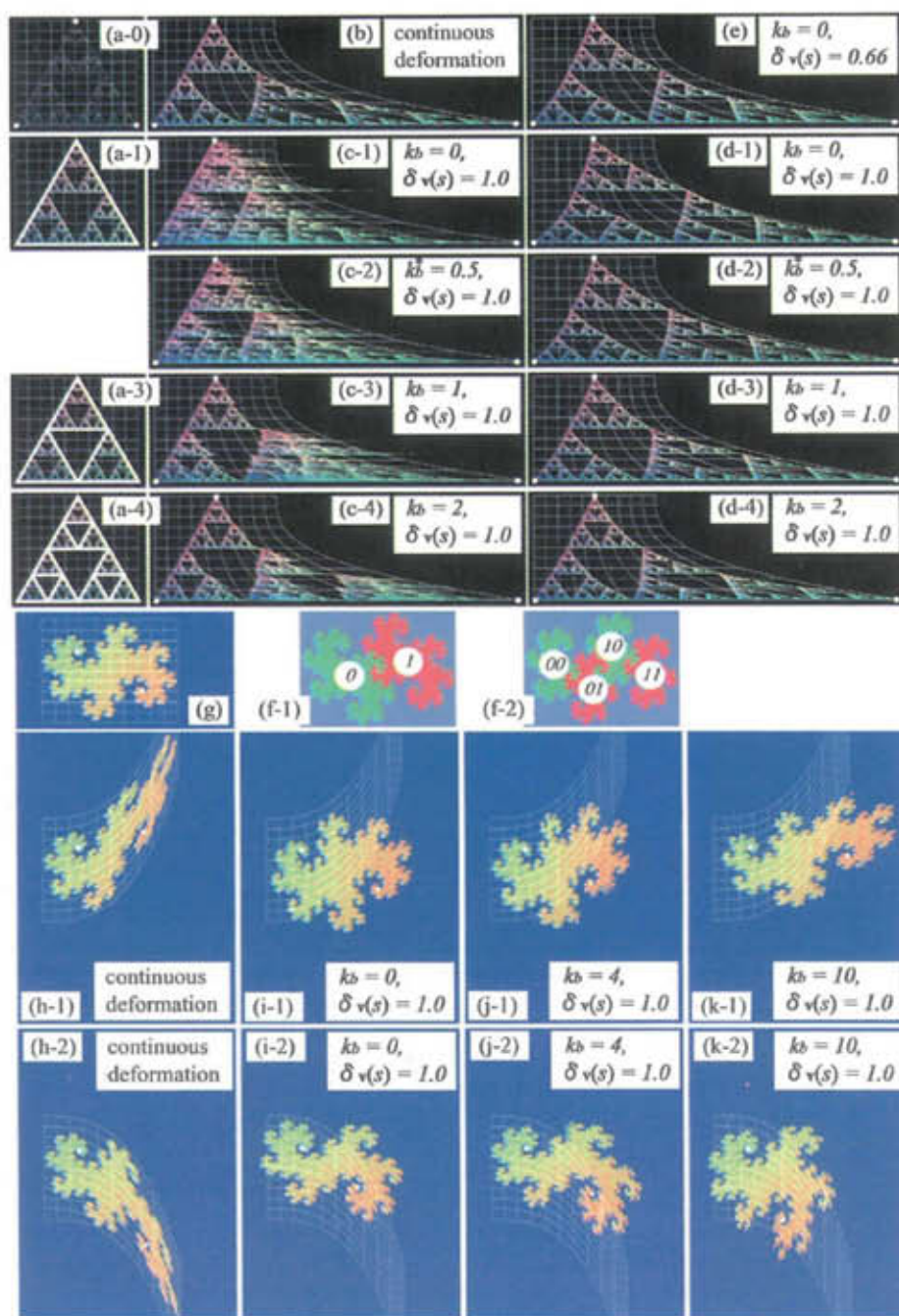


Figure 3. Fractal deformations (2). (a-0) — (e): Sierpinski gasket.  $a_i = d_i = 0.5, b_i = c_i = 0, i = 0, 1, 2, (e_0, f_0) = (-0.25, 0), (e_1, f_1) = (0.25, 0), (e_2, f_2) = (0, \sqrt{3}/4)$ . (f-1) — (k-2): Twin-dragon.  $a_i = b_i = d_i = 0.5, c_i = -0.5, i = 0, 1, (e_0, f_0) = (-0.5, 0.375), (e_1, f_1) = (0.5, 0.625)$ .



Comparing  $(i,j,k-1)$ , when  $k_b$  approaches  $k_e = 12$ , the deformation approaches  $(h-1)$ . The same applies to  $(i,j,k-2)$  and  $(h-2)$ .

The average calculation time for an image above is the following; f-squares: 6~8 min., Sierpinski gaskets: 1~3 min., Twin-dragons: 12~15 min.<sup>c</sup> We used Silicon Graphics 230 Visual Workstation 800 (Pentium III 800MHz, 768MByte).

## 5 Conclusion and Future Work

We have proposed a new fractal deformation technique by applying an ext-unit-IST to the increasing rates of d-vectors given to a geometric shape to deform, and shown its effectiveness. This technique provides easy control by changing a d-vector function intuitively, although real-time interactive operation is currently difficult. Improving its computational efficiency is required in future. This technique can be applied to any shape if it is given a proper address mapping. So, a method to give proper address mappings to various shapes other than IFS attractors is needed for practical use, such as deforming objects in pictures. Besides, we should extend this technique for shapes with multi-addressed points such as overlapping IFS's.

## References

1. Barnsley, M. F., *Fractals Everywhere*, 2nd ed., Academic Press, Boston, 1993.
2. Bowman, R. L., Fractal Metamorphosis: a Brief Student Tutorial, *Computers & Graphics*, Vol.19, No.1, pp.157-164, 1995.
3. Burch, B. and Hart, J. C., Linear Fractal Shape Interpolation, *Graphics Interface '97*, pp.155-162, 1997.
4. Fujimoto, T., Ohno, Y., Muraoka, K., and Chiba, N., Wrinkly Surface Generated on Irregular Mesh by Using IST Generalized on Code Space and Multi-Dimensional Space: Unification of Interpolation Surface and Fractal, *IEICE Transactions on Information and Systems*, Vol.E85-D, No.10, pp.1663-1677, 2002.
5. Fujimoto, T., Ohno, Y., Muraoka, K., and Chiba, N., Fractal Deformation Using Displacement Vectors Based on Extended Iterated Shuffle Transformation, *The Journal of the Society for Art and Science*, Vol.1, No.3, pp.134-146, 2002. <http://www.art-science.org/journal/v1n3/artsci-v1n3p134.pdf> .../v1n3/p134/index.html
6. Gomes, J., Darsa, L., Costa, B., and Velho, L., *Warping and Morphing of Graphical Objects*, Morgan Kaufmann, 1999.
7. Gonzalez, J. A., A Tutorial and Recipe for Moving Fractal Trees, *Computers & Graphics*, Vol.22, No.2-3, pp.301-305, 1998.
8. Massopust, P. R., *Fractal Functions, Fractal Surfaces, and Wavelets*, Academic Press, 1994.
9. Montiel, M. E., Aguado, A. S., and Zaluska, E. J., Topology in Fractals, *Chaos, Solitons and Fractals*, Vol.7, No.8, pp.1187-1207, 1996.
10. Sherman, P. and Hart, J. C., Direct manipulation of recurrent models, *Computers & Graphics*, Vol.27, No.1, pp.143-151, 2003.
11. Zair, C. E. and Tosan, E., Computer Aided Geometric Design with IFS Techniques, *Fractal Frontiers (Proc. Fractals '97)*, pp.443-452, 1997.

<sup>c</sup>For f-squares and Sierpinski gaskets, calculations were done for only the  $x$  direction; f-squares were given the same results for the  $y$  direction. For Twin-dragons, different calculations were done for the  $x$  and  $y$  directions respectively. We used conjugate gradient method to solve Eq.35.

# Multifractal and Stochastic Analysis of Electropolished Surfaces

MARIA HAASE, ALEJANDRO MORA AND BERND LEHLE<sup>1</sup>

*Institut für Computeranwendungen (ICA II),*

*Stuttgart University, Pfaffenwaldring 27, 70569 Stuttgart, Germany,*

<sup>1</sup>*vFlow Engineering GmbH, 70499 Stuttgart, Germany,*

*E-mail: mh@ica.uni-stuttgart.de, ica2am@csv.ica.uni-stuttgart.de, bernd@vflow.de*

Electropolishing is a century-old technical treatment used to obtain bright and shiny surfaces by electrochemical removal. Choosing the operating point in the transpassive region, the interplay between rising gas bubbles and a falling film of dissolved metal leads to complex surface structures. The choice of the electrolyte and the applied electrical potential have a significant influence on the surface structure on different length scales. The aim of the paper is a characterization of height profiles resulting from different electrolytes. Apart from estimating characteristic length scales and different scaling regions representing different physical or chemical processes, the multifractal scaling behaviour of the rough surfaces is investigated within the framework of wavelets. In particular, the wavelet transform modulus maxima method (WTMM) provides a robust estimation of the full spectrum  $D(h)$  of Hölder exponents  $h$  of the height profiles. From a statistical point of view the characterization of surface profiles by means of multifractal spectra is still incomplete. We apply a new stochastic approach, which is based on the theory of Markov processes and which allows the complete stochastic characterization of the surface profiles by means of a Kramers-Moyal expansion for the conditional probability distribution, which, in special cases, reduces to a Fokker-Planck equation describing the evolution of conditional probability distributions over scales.

**Keywords:** *Electropolishing, surface roughness, scaling, wavelet transform, maxima lines, Markov process*

## 1 Introduction

Electropolishing is a wide-spread technology discovered in the 1920s for obtaining smooth, shiny surfaces by electrochemical metal removal, i.e. the workpiece acts as an anode. The operating point of electropolishing is either the so-called *passive region*, where metal is dissolved slightly, or the *transpassive region*, where two competing processes occur in the case of vertically arranged electrodes: the dissolution of metal leading to a falling film of spent electrolyte containing dissolved metal, and the hydrolysis of water, where oxygen is formed at the anode causing gas bubbles to rise. The interaction of the two processes leads to local hydrodynamic instabilities, which mainly influence the formation of the surface structures on different scales<sup>1,2</sup>. As a result, an unwanted pattern of so-called gas lines appears on the micrometer scale<sup>3</sup>. The choice of the metal, the electrolyte and the applied electrical potential have a major influence on the surface structures.

In this study we investigate the surface topography of brass sheets, which have been electropolished in perpendicular position in the transpassive region. The workpieces are scanned with a 3D-laser-focus-scanner (UBM) with a micrometer resolution. Two different electrolyte solutions containing phosphoric acid and an amount of alcohol, denoted as METHANOL-electrolyte (METH) and GLYCERINE-electrolyte (GLYC), are used. Although metal surfaces electropolished in METH are very smooth, they show tiny ripples in the vertical direction caused by rising gas bubbles (Fig. 1a). In contrast, GLYC leads to a rougher surface structure with higher and

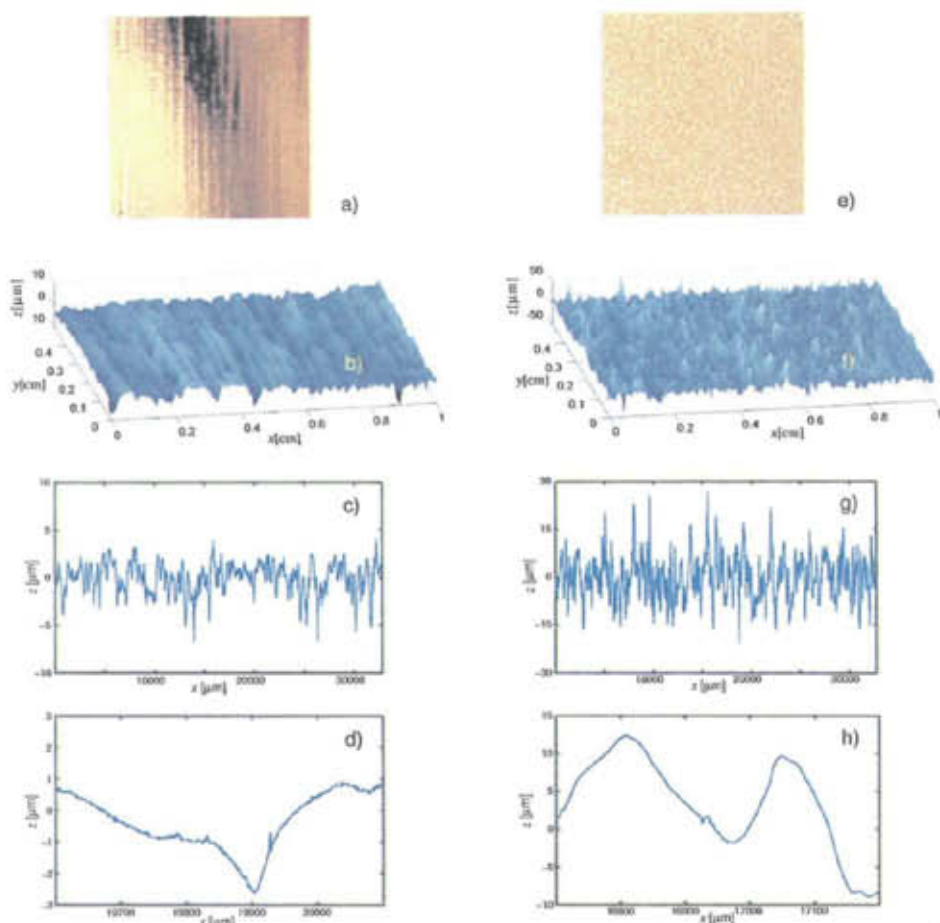


Figure 1: Brass surfaces electropolished in METH (a) and GLYC (e). Corresponding laser scans of height profiles  $z(x)$  and zooms transversal to the direction of gas lines for surfaces electropolished in METH (b,c,d) and GLYC (f,g,h).

sharper peaks causing a more mattfinished, isotropic appearance of the workpiece (Fig. 1b)<sup>1,2</sup>. We aim at a multifractal and stochastic characterization of surface height profiles depending upon the electrolyte used in order to get insight into the underlying interacting chemical and physical processes. This study is subdivided into three steps.

In the first step, classical spectral methods are used to extract the range of characteristic length scales introduced by the gas lines and different scaling regions indicating the interaction of different processes. While power spectra are useful in finding scaling regions, they only provide estimates about the *global* roughness or Hurst exponent  $h$  and thus about the self-affinity of the fractal height profile<sup>5,6</sup>. Information on the spatial distribution of possibly varying *local* Hurst or Hölder exponents, i.e. fluctuations in the surface roughness resulting from multi-affine properties, are

usually determined by means of the structure function method based on increments of the surface height profile<sup>7</sup>. However, there are fundamental limitations in the structure function approach, in particular, irregularities in the derivatives of the profiles can not be accessed. Using wavelets with more regularity and higher joint time-frequency resolution instead of increments (consisting of the difference of  $\delta$ -functions) allows for a complete and accurate multifractal analysis, which may even follow different power laws in different regions. The wavelet transform modulus maxima (WTMM) method<sup>8,9,13</sup> provides a robust method for the determination of singularity spectra<sup>10</sup>.

From the point of view of statistics, a characterization of the surface by means of multifractal spectra is still incomplete, since possible correlations of the roughness measures on different scales are not taken into consideration<sup>19,20,17</sup>. In a third step, we therefore apply a new stochastic approach based on the theory of Markov processes. This method allows to derive a stochastic differential equation for the evolution of the conditional probability density function (pdf) in the scale  $r$  directly from measurements without any assumption on the underlying data. The pdf's of increments on different length scales display a similar deformation as in the case of turbulent flow indicating intermittency effects. The analysis of multiconditional pdf's suggests that the statistics has Markov properties, which is the pre-condition for the applicability of the new stochastic method<sup>20</sup>.

The paper is organized as follows. In section 2, we estimate the range of length scales characterizing the gas lines and different scaling regions. After a brief review of the continuous wavelet transform (CWT) and the wavelet transform modulus maxima (WTMM) method, the multifractal spectra for surfaces electropolished in METH and GLYC, respectively, are estimated in section 3. It can be seen that at least two processes are interacting leading to different multifractal behaviour on different scales. Section 4 deals with the evolution of the probability density distributions of surface height increments for varying scales and gives a short introduction into the recently developed stochastic approach based on the theory of Markov processes, which is used for a complete stochastic characterization of the profiles. A preliminary stochastic analysis of the height profiles based on the theory of Markov processes is given. Finally, section 5 presents our conclusions and perspectives of further investigations.

## 2 Characteristic length scales and scaling regions

The laser scan (fig. 1b) of a brass surface electropolished in METH (fig. 1a) shows, that the gas lines introduce a natural length scale into the surface structure. On the other hand, the brass sheet electropolished in GLYC (fig. 1e,f) has a rough, fractal-like structure with high peaks, suggesting scaling properties. Figs. 1c,d,g,h show typical surface height profiles transversal to the direction of the gas lines together with zooms for METH and GLYC. Note the different scales in  $z$ -direction.

In order to estimate the characteristic length scale introduced by the gaslines we measured the anisotropy introduced by the rising bubbles calculating the power spectral density of ensemble averages transversal and parallel to the gas lines for METH (figs. 2a,b). From the ratio of the corresponding power spectra, the characteristic length scales can be readily extracted (about 24 gas lines/cm)<sup>4</sup>.

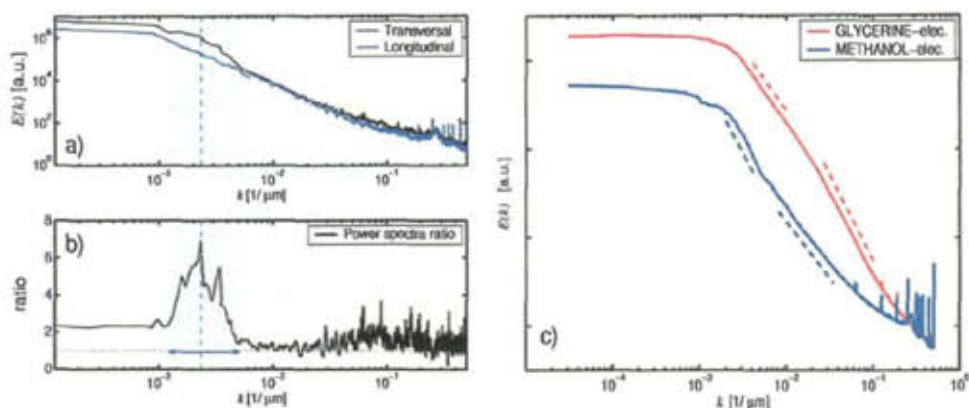


Figure 2: Comparison of power spectra of ensemble averages transversal and parallel to the gas lines for METH (a). From the ratio of power spectral densities (b), the range of length scales corresponding to the gas lines can be extracted. Power spectral densities  $E(k)$  for ensemble averages of surface profiles transversal to the gas lines for METH and GLYC (c).

Fig. 2c shows the power spectral density  $E(k)$  as a function of the wave number  $k$  for METH and GLYC, where we considered ensemble averages of profiles transversal to the direction of the gas lines and applied moving window averaging. For both power spectra, two regions with different decay can be seen, indicating the interplay of two different processes. For large scales the workpieces electropolished in METH appear much smoother than those treated with GLYC. Therefore, the METH power spectrum displays a faster decay for large scales than the GLYC power spectrum. For small scales the behaviour is just reversed (cf. fig.1).

### 3 Multifractal analysis based on wavelets

While the power spectral density  $E(k)$  is of great interest revealing scaling laws, it gives only limited information about the mono- or multifractal properties of the surface roughness. It only allows for an estimation of a *global* Hurst or Hölder exponent  $h$  via the relation  $E(k) \sim k^{-1-2h}$ . Local fluctuations in the degree of 'roughness' call for a location-dependent Hölder exponent  $h(x)$ . The standard way to extract the multiscaling properties of a function  $f(x)$ , is to study the scaling behaviour of the structure functions  $S_q(r)$ <sup>7</sup>

$$S_q(r) = \langle \delta f_r^q \rangle \sim r^{\zeta_q} \quad (1)$$

of order  $q$  of the increments  $\delta f_r = f(x+r/2) - f(x-r/2)$ . Multifractal behaviour leads to a nonlinear scaling exponent  $\zeta_q$ . By Legendre transforming the exponents  $\zeta_q$ , one obtains an estimate of the spectrum  $D(h)$  of Hölder exponents<sup>7</sup>

$$D(h) = \min_q (qh - \zeta_q + 1). \quad (2)$$

A severe drawback of this method is, that one has only access to Hölder exponents  $0 < h < 1$ , i.e. singularities in the derivatives of the function can not be identified. In addition, negative moments  $q < 0$  lead to divergencies<sup>9</sup>.



These limitations can be circumvented using the wavelet framework. The continuous wavelet transform (CWT)

$$W_\psi f(a, b) = \frac{1}{a} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx \quad (a, b \in \mathbf{R}, a > 0) \quad (3)$$

decomposes the function  $f(x) \in L^2(\mathbf{R})$  hierarchically in terms of elementary components  $\psi\left(\frac{x-b}{a}\right)$  which are obtained from a single *mother wavelet*  $\psi(x)$  by dilations and translations. Here,  $a$  denotes the scale and  $b$  the shift parameter. A unique reconstruction of the function  $f(x)$  is ensured if  $\psi(x) \in L^1(\mathbf{R})$  has zero mean.

Increments can be rewritten as *poor man's* wavelets consisting of the difference of two  $\delta$  functions<sup>9</sup>. Choosing wavelet functions, which are well localized both in physical and Fourier space a much better joint space/frequency resolution can be achieved leading to an improved data analysis. Among the multitude of possible choices for wavelets the Gaussian family of real wavelets, which are obtained as derivatives of the Gaussian function,

$$\psi_0(x) = e^{-x^2/2}, \quad \psi_n(x) = \frac{d}{dx} \psi_{n-1}(x) \quad (n \in \mathbf{N}, n \geq 1). \quad (4)$$

is especially suitable for detecting and characterizing irregularities in a function or even in its derivatives. For this purpose, we require wavelets  $\psi(x)$  with  $n_\psi$  vanishing moments

$$\int_{-\infty}^{+\infty} x^k \psi(x) dx = 0, \quad \forall k, \quad 0 \leq k < n_\psi. \quad (5)$$

Wavelets allow to precisely detect and quantify singularities. Assuming a cusp singularity with Hölder exponent  $h(x_0) \in (n, n+1)$  at  $x_0$ , the CWT scales like

$$|W_\psi f(a, x_0)| \sim a^{h(x_0)}, \quad a \rightarrow 0^+, \quad (6)$$

provided the analyzing wavelet chosen has  $n_\psi > h(x_0)$  vanishing moments<sup>8,9</sup>. In contrast, if one chooses a wavelet with  $n_\psi < h(x_0)$ , the CWT scales with an exponent  $n_\psi$ . It can be shown, that this scaling behaviour is also valid along the maxima lines of the modulus of the CWT, which point to the singularities<sup>8</sup>. Hence, for practical applications, these lines are conveniently used for extracting the Hölder exponents and can be regarded as *fingerprints* containing the complete information on the scaling behaviour. A direct tracing of these maxima lines reduces the time-consuming calculation of the full redundant CWT even so being sufficient for the characterization of the scaling properties of  $f(x)$ <sup>13</sup>. The so-called wavelet transform modulus maxima (WTMM) method<sup>9,10</sup> allows (in the absence of oscillating singularities) a robust estimation of the full spectrum of singularities.

The WTMM method is a generalization of the classical multifractal formalism<sup>11,12</sup>, where box functions are replaced by wavelets as oscillating variants. The partition function  $Z(q, a)$  can be considered as a modified wavelet based structure function<sup>9</sup>

$$Z(q, a) = \sum_{b_i \in \text{max. lines}} \left( \sup_{a' \leq a} |W_\psi f(a', b_i)| \right)^q \quad (7)$$

For a given scale  $a$ ,  $Z(q, a)$  contains the  $q$ th moments of the contributions of  $|W_\psi f|$  along the maximal lines, where the supremum in eq. (7) is related to a Hausdorff-like covering with scale-adapted wavelets. Using this definition, divergencies due to negative order moments are removed<sup>9</sup>. From the power-law behaviour of the partition function (cf. eq.1),  $Z(q, a) \sim a^{\tau(q)}$ ,  $a \rightarrow 0^+$ , the whole spectrum of Hölder exponents  $D(h)$  is obtained by Legendre transforming the scaling exponents  $\tau(q)$ :

$$D(h) = \min_q (qh - \tau(q)). \quad (8)$$

The spectrum of Hölder exponents  $D(h)$  is used to characterize the fluctuations in the roughness of electropolished surfaces. We apply the WTMM method to profiles transversal to the gas lines for brass sheets electropolished in METH and GLYC. For each electrolyte, we use an ensemble of five profiles transversal to the direction of the gas lines for the calculation of the partition function  $Z(q, a)$  defined in eq. (7). In accordance with the observation of two different scaling regions in the power spectra (fig. 2c), two regions with different power law behaviour occur, which lead to different distributions of the corresponding Hölder exponents. Up to 12800 maxima lines are used for the evaluation of  $D(h)$ . The corresponding singularity spectra  $D(h)$  for small and large scales are plotted in fig. 3. Due to the finite resolution of measurements and cross-over effects between small and large scales, the spectra depend to some extent on the scale interval selected for the evaluation of  $D(h)$ .

The spectra of Hölder exponents displayed in figs. 3a,b show that the strengths of singularities for METH and GLYC are quite similar on small scales. To be more specific: although the support of the  $D(h)$  curve of METH is slightly shifted to smaller values of  $h$  as compared to the  $D(h)$  curve of GLYC, the most frequent Hölder exponent  $h_0$  is around 1.25 in both cases. A further difference is that the maximum  $D(h_0)$  is smaller for GLYC indicating that the corresponding singularities are less frequent. Thus, in contrast to the visual impression received from figs. 1b,c, surfaces electropolished in GLYC are smoother on small scales than those treated with METH (cf. fig. 2c). For large scales, however, the GLYC spectrum is clearly shifted to smaller  $h$ -values with  $h_0 = 0.79$  as compared to  $h_0 = 1.02$  for METH, which corroborates the observation, that GLYC produces sharper higher peaks on large scales than METH. The spectra  $D(h)$  can therefore be considered as a quantification of the qualitative conclusions drawn from inspecting the profiles shown in fig. 1. The main difference in the surface topography of brass workpieces electropolished in different electrolytes seems to be caused by the rising gas bubbles, their magnitude, adhesion time and dynamics producing different patterns on large scales. This is due to the fact that the gas bubbles are generating a flow of fresh electrolyte etching the surface in the gas lines effectively<sup>1,4</sup>.

#### 4 Stochastic Analysis

For rough surfaces displaying scaling behaviour the singularity spectrum  $D(h)$  can be regarded as a complete multifractal characterization of the singularities<sup>5,6</sup>. However, from a stochastic point of view, this characterization is still incomplete, since

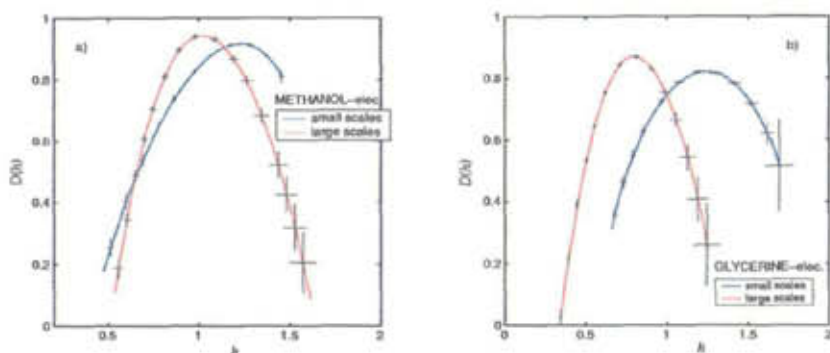


Figure 3: Spectra of Hölder exponents  $D(h)$  for (a) METH and (b) GLYC for different scaling regions.

joint statistical properties of several height increments on different scales are not taken into account<sup>17,18</sup>. In a series of papers a new approach for the stochastic analysis has been proposed which allows to extract the explicit form of the underlying stochastic process directly from experimentally measured data without making any assumptions, provided the process is Markovian<sup>19,20,17</sup>. Considering the height increment  $z_r(x) = z(x + r/2) - z(x - r/2)$  of a surface profile  $z(x)$  as a stochastic variable in the length scale  $r$ , the aim is to describe the evolution of the conditional probability density function (pdf) as  $r$  is varied, where the conditional pdf  $p(z_1, r_1 | z_2, r_2)$  describes the probability for finding the increment  $z_1$  on scale  $r_1$  provided that the increment  $z_2$  is given on scale  $r_2$ . A stochastic process is Markovian, if the conditional probability densities fulfil the relations

$$p(z_1, r_1 | z_2, r_2; \dots; z_n, r_n) = p(z_1, r_1 | z_2, r_2) \quad \text{where} \quad r_1 < r_2 < \dots < r_n. \quad (9)$$

In this case, the conditional pdf satisfies a master equation. Expanding the distribution function into a Taylor series, the evolution equation can be written as<sup>21,20</sup>

$$-r \frac{\partial}{\partial r} p(z_r, r | z_0, r_0) = \sum_{k=1}^{\infty} \left( -\frac{\partial}{\partial z_r} \right)^k D_k(z_r, r) p(z_r, r | z_0, r_0) \quad (10)$$

where the so-called Kramers-Moyal coefficients  $D_k(z_r, r)$  can be directly estimated from experimental data

$$D_k(z_r, r) = \lim_{\Delta r \rightarrow 0} M_k(z_r, r, \Delta r), \quad \text{where} \quad (11)$$

$$M_k(z_r, r, \Delta r) = \frac{r}{k! \Delta r} \int_{-\infty}^{\infty} (\tilde{z} - z_r)^k p(\tilde{z}, r - \Delta r | z_r, r) d\tilde{z} \quad (12)$$

In the special case of  $D_4(z_r, r) = 0$ , eq. (10) reduces to the Fokker-Planck equation

$$-r \frac{\partial}{\partial r} p(z_r, r | z_0, r_0) = \left\{ -\frac{\partial}{\partial z_r} D_1(z_r, r) + \frac{\partial^2}{\partial z_r^2} D_2(z_r, r) \right\} p(z_r, r | z_0, r_0). \quad (13)$$

The partial differential equations (eqs. 10,13) completely describe the underlying stochastic process. For details the reader is referred to<sup>21,20</sup>.



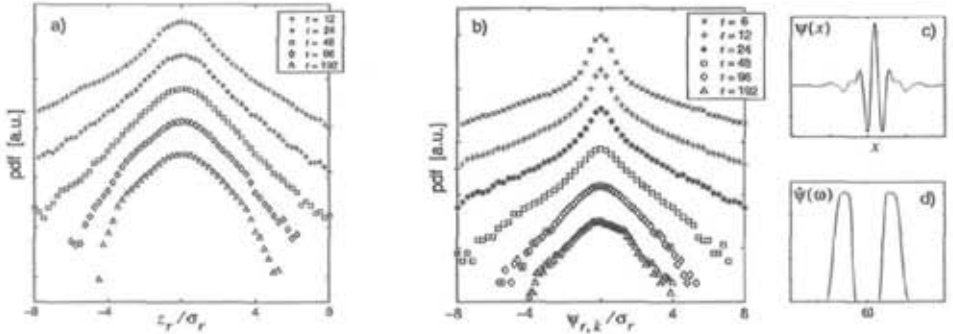


Figure 4: Probability density functions of for METH based on (a) height increments and (b) Meyer wavelet coefficients for different scales  $r_i$ . Meyer wavelet (c) and its Fourier transform (d).

Next, we show first results needed for the application of this stochastic approach. In figs. 4a,b the probability density functions of  $z$ -increments and Meyer wavelet coefficients<sup>16</sup> (see figs. 4c,d) for METH are plotted for various scales  $r_i$ . The pdf's are normalized to their respective standard deviations  $\sigma_r$  and shifted in vertical direction for clarity. For small scales the shapes of the curves deviate strongly from Gaussian distributions indicating pronounced intermittency effects. In a next step, we test the data for evidence of an underlying Markovian process. Since a general

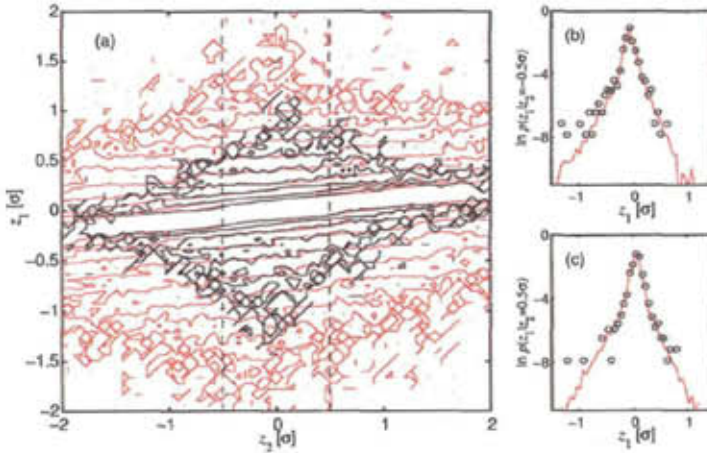


Figure 5: (a) Contour lines of the conditional pdf's  $p(z_1, r_1 | z_2, r_2)$  (red lines) and  $p(z_1, r_1 | z_2, r_2; z_3 = 0, r_3)$  (black lines) for  $r_1 = 10 \mu\text{m}$ ,  $r_2 = 108 \mu\text{m}$ ,  $r_3 = 216 \mu\text{m}$  for METH (b), (c) Cuts through the conditional pdf's for  $z_2 = \pm \sigma/2$ .

test of the condition, eq. 9, for all sets of scales  $r_1, r_2, \dots, r_n$  and for all  $n$  is not possible we test the validity of the following necessary condition

$$p(z_1, r_1 | z_2, r_2; z_3, r_3) = p(z_1, r_1 | z_2, r_2) \quad \text{where} \quad r_1 < r_2 < r_3. \quad (14)$$

The results for a brass sheet electropolished in METH are presented in figs. 5,6. In Fig. 5a, the contour plots of  $p(z_1, r_1 | z_2, r_2; z_3, r_3)$  (black lines) and  $p(z_1, r_1 | z_2, r_2)$  (red lines) are shown in units of the standard deviation  $\sigma$  of the  $z$ -data. The good

correspondence over several orders of magnitude is corroborated by two cuts for  $z_2 = \pm\sigma/2$  displayed in figs. 5b,c indicating the validity of the necessary condition eq. (14). In fig. 6a the same contour plots are presented for a different choice of scale increments. In this case, the two sets of contour lines deviate strongly from each other, i.e. here a Markovian condition is not satisfied. Similar results are obtained for the GLYC case. Thus, the Markov properties are likely to hold for large enough differences in the scales  $r_i$ , but are violated for small differences. Further investigations are necessary to estimate the range of validity of Markov properties. In a next step, we estimate the Kramers-Moyal coefficients  $D_1, D_2$  and  $D_4$  (eqs. 11,12), which allow to set up eq. (10) or eq. (13), respectively, and thus to describe the stochastic process completely.

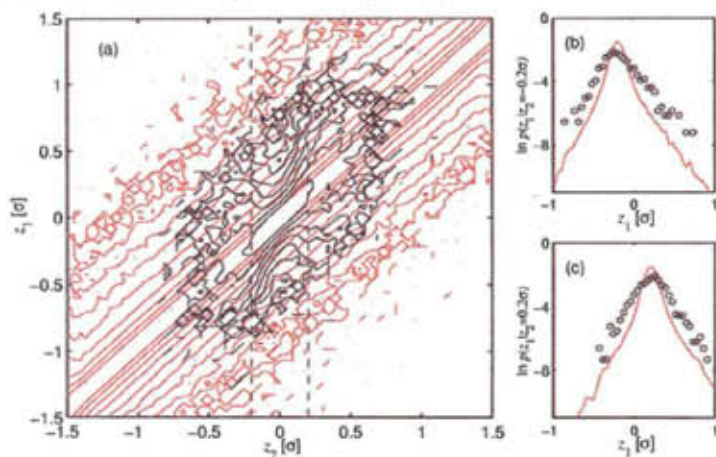


Figure 6: Like figure 5, but for different scale increments:  $r_1 = 52 \mu\text{m}$ ,  $r_2 = 60 \mu\text{m}$ ,  $r_3 = 68 \mu\text{m}$ .

## 5 Conclusions

We presented various numerical techniques including the wavelet framework and stochastic methods for a characterization of surface electropolished profiles. The complete information about the multifractal scaling behaviour is contained in the singularity spectra  $D(h)$  which are estimated using the WTTM method. We have shown, that it is possible to distinguish between the surface structures caused by two interacting processes, namely rising gas bubbles introducing an anisotropy into the topography, and the dissolution of metal. This anisotropy is more pronounced in the case of METH. While workpieces electropolished in GLYC appear rougher on a large scale than those treated with METH, an opposite tendency is observed on the small scales. The main difference in the surface structures is caused by the specific peculiarity of the gas lines. For a complete stochastic description of the surfaces, we use a new approach based on the theory of Markov processes.

Further investigations include the question, if the Markov analysis allows to set up a Fokker-Planck equation for the different interacting processes for electropolishing in the transpassive regime. The corresponding Langevin equation would open the possibility for a direct synthetization of surface profiles. Also, we are interested in whether or not one could benefit for the Markov analysis from a replacement

of the height increments by wavelet coefficients or the CWT-values of  $W_\psi z(a, b_i)$  along the maxima lines, respectively. This work is under progress.

We gratefully acknowledge the supply of surface measurements by C. Gerlach and many stimulating discussions with R. Friedrich, J. Peinke, M. Wächter, A. Kouzmitchev and F. Kun. This work was supported by the VolkswagenStiftung grant I/77315.

## References

1. C. Gerlach, *PhD thesis*, University of Bremen (2002).
2. C. Gerlach, A. Visser, P.J. Plath, in: *Nonlinear Dynamics of Production Systems*, (G. Radons, R. Neugebauer eds.) Wiley-VCH, Weinheim (2003).
3. M. Buhler, *Fortschritt-Berichte VDI*, No.553, VDI-Verlag Düsseldorf (2000).
4. A. Mora, C. Gerlach, T. Rabbow, P.J. Plath, M. Haase, in: *Nonlinear Dynamics of Production Systems*, (G. Radons, R. Neugebauer eds.) Wiley-VCH, Weinheim (2003).
5. J. Feder, *Fractals*, Plenum Press, New York (1988).
6. A.-L. Barabási, H.E. Stanley, *Fractal concepts in surface growth*, Cambridge University Press (1995).
7. U. Frisch, G. Parisi, in *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, (M.Ghil et al. eds.) North Holland, Amsterdam, 71–88 (1985).
8. S. Mallat and W.L. Hwang, *IEEE Trans. Inf. Theory*. **38**, 617–643 (1992).
9. J.F. Muzy, E. Bacry and A. Arnéodo, *Int. J. Bif. Chaos* **4**, 245–302 (1994).
10. S. Jaffard, in *Fractals in Engineering*, (J. Lévy Véhel et al. eds.), Springer, London, 2–18 (1997).
11. T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia and B.I. Shraiman, *Phys. Rev. A* **33**, 1141–1151 (1986).
12. J. Argyris, G. Faust and M. Haase, *An Exploration of Chaos*, North-Holland, Amsterdam (1994).
13. M. Haase, B. Lehle, in: *Fractals and Beyond*, (M. M. Novak ed.), World Scientific, Singapore 241–250 (1998).
14. A. Arnéodo, E. Bacry, S. Jaffard, J.F. Muzy, *J. Stat. Phys.* **87**, 179–209 (1997).
15. M. Haase, in: *Paradigms of Complexity*, (M. M. Novak ed.), World Scientific, Singapore 287–288 (2000).
16. I. Daubechies, *Ten lectures on wavelets*, SIAM, Philadelphia, PA (1992).
17. M. Wächter, F. Riess, H. Kantz, J. Peinke, Stochastic analysis of surface roughness, arXiv:physics/020368 and submitted to *Europhys. Lett.*
18. C. Renner, J. Peinke, R. Friedrich, O. Chanal, B. Chabaud, *Phys. Rev. Lett.* **89**, 124502 (2002).
19. R. Friedrich, J. Peinke, *Physica D* **102**, 147–155 (1997).
20. C. Renner, J. Peinke, R. Friedrich, *J. Fluid Mech.* **433**, 383–409 (2001).
21. H. Risken, *The Fokker-Planck Equation*, Springer Berlin (1996).

# A METHOD FOR NUMERICAL ESTIMATION OF GENERALIZED RÉNYI DIMENSIONS OF AFFINE RECURRENT IFS INVARIANT MEASURES

TOMEK MARTYN

*Computer Graphics Laboratory, Institute of Computer Science, Warsaw University of  
Technology, ul. Nowowiejska 15/19, 00-665, Warsaw, Poland  
E-mail: martyn@ii.pw.edu.pl*

In this paper we address and propose a solution to the problem of numerical estimating the generalized  $q$ -dimensions of affine RIFS invariant measures. Unlike the commonly used chaos game approach, our method gives good results for the potentially whole range of  $q$  (including the problematic large negative  $q$ ) in an efficient and robust manner. In this goal, we use a deterministic, Markov-operator-based algorithm of approximating the measure on a lattice. We show that the algorithm makes it possible to approximate the measure at any accuracy with respect to the Hutchinson metric. Then, we give the rigorous proof that our lattice approximation is ideally suit for computing the generalized dimensions by means of the overlapping box approach. The results included and their comparison with those obtained from the chaos game confirm the strength of our approach when applied in practice.

## 1 Introduction

Originated by the Hungarian mathematician A. Rényi in information theory and introduced to fractal geometry in <sup>1</sup>, the generalized dimensions are one of the basic tools to describe global properties of fractal distributions (measures) in the form of powerlaw behavior of  $q$ th order moment generating functions. First of all, thanks to the whole range of values that the generalized dimensions take in the case of a multifractal measure, they can be considered one of the ways of quantitative characterization of *multifractality* and, thus, the most fundamental property distinguishing multifractals from monofractals, because for the latter all the dimensions are just equal to the box-counting dimensions of the supports. Secondly, via the Legendre transform, they give the basis for describing local scaling behavior of a measure in the form of the multifractal spectra and, further, for multifractal analysis in general. Therefore, there is no surprise that much work has been devoted to find explicit formulas for the dimensions of different classes of measures. As a result, the formulas are known for self-similar measures: from the classical binomial and multinomial measures to infinite multinomial ones<sup>2</sup>, and further for self-similar measures defined by the multiplicative Moran cascade processes<sup>3</sup> and their extensions to graph-directed constructions<sup>4</sup>, as well as for random analogues of the above classes<sup>5,6,7</sup>. While the self-similar class is well-investigated, the self-affine measures are much harder to analyze and only in rather special cases the explicit formulas for the generalized dimensions can be derived<sup>8,9</sup>. In addition, all the mentioned classes require the open set condition to hold (more precisely, in <sup>7</sup> a little bit weaker condition was assumed). In the case of the general self-affine measures arising from, possibly, overlapping constructions the only way to deal with multifractal analysis in general and the generalized dimensions in particular is via numerical estimation.



The common way to analyze measures arising from general IFS constructions is to approximate an IFS measure on a lattice from a *long* trajectory of points generated by the chaos game<sup>10</sup> and, then, apply lattice-based (i.e. box-counting) methods (see e.g.<sup>11</sup>). By Elton's ergodic theorem<sup>12</sup> the time-averages along the random process converge to an invariant measure (which in the IFS case is unique), so this dynamical-system approach seems to be, at least theoretically, legitimate. But is it in practice? As far as the subject of interest are the generalized dimensions the answer is, unfortunately, 'no' in general, and this follows from the vague "long" concerning the chaos game trajectory. First of all, there are no error bounds saying how long the trajectory should be so as to approximate an IFS measure at accuracy given. What is more important, however, is that probability of reaching a box of a positive measure by a finite trajectory can be arbitrarily small (just by treating an IFS measure in the "residence-time" manner). Consequently, the boxes of a nonzero but small measure relatively to the others are either not visited or the number of the chaos game steps would have to be arbitrarily large in general. Thereby, the near-zero-measure boxes are undistinguishable from the zero-measure ones. While the omitted boxes have almost no impact on the values of the generalized  $q$ -dimensions for positive  $q$  (because their contribution to the moment generating function is almost null), the case of negative  $q$  reverses this situation and the boxes of small measure play now the major role in calculating the dimensions. As a result, the practical application of the chaos game approach usually significantly underestimates the generalized dimensions for large negative  $q$ .

In this paper we cure this problem by exploiting the other, i.e. space-mean, side of ergodicity and approximate a RIFS measure on a lattice in a deterministic way, using the measure's invariance under the corresponding Markov operator. It should be noted that a totally different deterministic method of approximating a "usual" IFS measure (which also utilizes the definition of the Markov operator) has recently been proposed in<sup>13</sup>. Though that method is theoretically appealing, yet given  $n \times n$  it requires computing the left-eigenvector of a  $n^2 \times n^2$  transition matrix of a Markov chain, which makes it both memory and computationally expensive (ignoring the problems related to numerical stability). Our method of approximation is based on the ideas used previously in<sup>14</sup> (which we adapt here to the lattice context) and, in contrast to<sup>13</sup>, is efficient, numerically stable, and extends the scope to Recurrent IFS measures. The algorithm is completed with the proof concerning the quality of approximation it provides. Then, we give the rigorous proof that our method when accompanied with the overlapping box approach<sup>15</sup> is ideally suit for the numerical estimation of the generalized dimensions. Finally, we compare results supplied by our approach with those obtained from the chaos game.

## 2 Preliminaries

### 2.1 Generalized Rényi dimensions

Let  $\mu$  be a Borel measure on  $\mathbb{R}^d$  with a bounded support. Let  $\delta_n$  be a sequence such that  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ . Denote by  $G_{\delta_n}$  the collection of boxes  $B = \prod_{i=1}^d [a_i \delta_n, (a_i + 1) \delta_n)$ ,  $a_i \in \mathbb{Z}$ , belonging to the  $\delta_n$ -coordinate lattice

covering of  $\mathbb{R}^d$ , so that  $\mu(B) > 0$ . Define the function  $U_{\delta_n}: G_{\delta_n} \rightarrow \mathbb{R}$  by  $U_{\delta_n}(B) = -\log \mu(B)$ . Typically, the *rescaled moment generating function* is a renormalized version of the Laplace transform  $\mathbb{E}[\exp(-\lambda U_{\delta_n})]$  of  $U_{\delta_n}$ , and is defined by  $\tau(q) := -\lim_{n \rightarrow \infty} \log_{\delta_n} \mathbb{E}[\exp(-q U_{\delta_n})]$  if the limit exists, where  $q \in \mathbb{R}$  and the expectation is taken with respect to the Lebesgue measure. After rewriting,  $\tau$  takes the well-known form

$$\tau(q) = \lim_{n \rightarrow \infty} \frac{\log \sum_{B \in G_{\delta_n}} \mu(B)^q}{-\log \delta_n}. \quad (1)$$

The *generalized Rényi (lattice) dimensions*  $D(q)$  are defined by

$$D(q) := \begin{cases} \frac{\tau(q)}{1-q} & q \neq 1, \\ \lim_{n \rightarrow \infty} \frac{\sum_{B \in G_{\delta_n}} \mu(B) U_{\delta_n}(B)}{-\log \delta_n} & \text{otherwise.} \end{cases} \quad (2)$$

Unfortunately, there are some problems with (1) (and, hence, with (2)) even if it is considered from the pure theoretical viewpoint<sup>15</sup>. First, the limit sometimes depends on the sequence  $\delta_n$ , which is not only undesired theoretically but also makes numerical estimation unstable. Secondly, just like the Laplace transform, the limit is often not well-defined for the negative values of  $q$ —more precisely it is just  $\infty$  whenever  $q < 0$ . As shown in <sup>15</sup>, a method, which improves the behavior of  $\tau$ , is to use the lattice of overlapping boxes  $(B)_\kappa = \prod_{i=1}^d [(a_i - \kappa)\delta_n, (a_i + 1 + \kappa)\delta_n]$ , i.e. each box  $B$  of the standard lattice is "blown" up by a factor  $\kappa \in \mathbb{N} = \{1, 2, \dots\}$ .

Let us denote by  $f^\# \nu$  the image of a measure  $\mu$  on  $X$  under the mapping  $f: X \rightarrow X$ , that is  $(f^\# \nu)(E) := \nu(f^{-1}(E))$  for  $E \subset X$ . The following proposition proved in <sup>15</sup> states an important result about the overlapping box approach, which we will use in the sequel.

**Proposition 1.** *Let  $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a bi-Lipschitz map. Let  $(\delta_n)$  be a sequence such that  $\delta_n \geq \delta_{n+1} \geq c\delta_n$  for all  $n \in \mathbb{N}$  and a constant  $c > 0$ . Then for all  $q \in \mathbb{R}$*

$$\limsup_{n \rightarrow \infty} \frac{\log \sum_{B \in G_{\delta_n}} \mu((B)_{\kappa_1})^q}{-\log \delta_n} = \limsup_{\delta \downarrow 0} \frac{\log \sum_{B \in G_\delta} (f^\# \mu((B)_{\kappa_2}))^q}{-\log \delta} \quad (3)$$

for any  $\kappa_1, \kappa_2 \in \mathbb{N}$ .<sup>a</sup>

## 2.2 IFS and RIFS invariant measures

Let  $K$  be a compact metric space. A *recurrent iterated function system* (abbreviated RIFS) is a triple  $(K; \{w_i\}_{i=1}^N; P)$ , where  $w_i: K \rightarrow K$  are contractive maps and  $P \in \mathbb{R}^{N \times N}$  is an irreducible row-stochastic matrix,  $N \in \mathbb{N}$ . From the viewpoint of Markov chain theory,  $P = [p_{ik}]$  is a transition probability matrix which specifies a homogenous Markov chain on  $N$  states  $\{w_i\}$ , so that probability of transition from  $w_i$  to  $w_k$  is given by  $p_{ik}$ . Since a homogenous irreducible Markov chain with finite state space is positive recurrent, it follows that the chain is ergodic with the unique stationary distribution  $\pi = [\pi_1, \dots, \pi_N]$  being the unique normalized left eigenvector of  $P$ , i.e.,  $\pi = \pi P$  and  $\sum_{i=1}^N \pi_i = 1$ . (If, for each  $k \in \{1, \dots, N\}$ ,

<sup>a</sup>Note that in (3) the sums run over the boxes  $B$  with  $\mu(B) > 0$ , and not the expanded  $(B)_\kappa$ .

$p_{ik} = p_k$  for every  $i \in \{1, \dots, N\}$ , then the RIFS reduces to the "ordinary" IFS written usually as  $(K; \{w_i\}_{i=1}^N; \{p_i\}_{i=1}^N)$ . Evidently, in such a case  $\pi = [p_i]_{i=1}^N$ .

Each RIFS is the input for the so-called *chaos game* which, exploiting the underlying Markov chain given by  $P$ , generates a random walk  $\{x_i\}_{i=0}^\infty$  on  $K$ , so that  $x_n = w_{i_n}(x_{n-1})$  and  $x_0$  is any point of  $K$ . One can show that the walk is a random process of asymptotic behavior specified by a unique stationary (Borel) measure<sup>16</sup>  $\mu_r(E) = \lim_{n \rightarrow \infty} \frac{1}{n+1} \# \{ \{x_i\}_{i=0}^n \cap E \}$ , whenever  $E$  is a Borel set with  $\mu_r(\partial E) = 0$ . Moreover,  $\mu_r$  is supported by a set  $A$  being the unique compact and nonempty solution of the set equation

$$A = \bigcup_{i=1}^N A^{(i)}, \quad A^{(i)} = \bigcup_{\{k: p_{ki} > 0\}} w_i(A^{(k)}), \quad (4)$$

that is  $\text{spt } \mu_r = A$  is the *attractor* of the RIFS.

The main conclusion of Elton's ergodic theorem<sup>12,16</sup> is that the time mean  $\mu_r$  coincides (almost everywhere) with the space mean  $\int_X \mathbf{1}_E d\mu$  in which  $E$  is Borel, and  $\mu$  is the  $K$ -projection of a unique measure being invariant for the "vector" version of the so-called *Markov operator*. (By  $\mathbf{1}_E$  we denote the indicator function of a set  $E$ .) More precisely, the Markov operator associated with a RIFS  $(K; \{w_i\}_{i=1}^N; P)$  acts on the space  $\mathcal{P}(K^N)$  of normalized Borel measures on the Cartesian  $N$ -product of  $K$ , and is defined by

$$\widetilde{M}(\tilde{\nu}) := \left[ \sum_{k=1}^N p_{k1} w_1^\# \nu_k, \dots, \sum_{k=1}^N p_{kN} w_N^\# \nu_k \right], \quad (5)$$

where  $\tilde{\nu} = [\nu_1, \dots, \nu_N]$ ,  $\nu_i$  are Borel measures on  $K$  such that  $\sum_{i=1}^N \nu_i(K) = 1$ . One can show that  $\widetilde{M}$  possesses a unique fixed point  $\tilde{\mu} = \widetilde{M}(\tilde{\mu})$  on the subspace  $\{\tilde{\nu} \in \mathcal{P}(K^N) : \nu_i(K) = \pi_i, i = 1, \dots, N\}$ , where  $\pi$  is the unique distribution of the underlying Markov chain. Moreover,  $\text{spt } \tilde{\mu} = [A^{(1)}, \dots, A^{(N)}]$ . Thus, the RIFS invariant measure can be expressed in terms of  $\tilde{\mu}$  as

$$\mu(E) \equiv \tilde{\mu}(E^N) = \sum_{i=1}^N \mu_i(E) \quad \text{for } E \subset K. \quad (6)$$

Both from the theoretical and practical viewpoint, it is often useful to introduce a function that says how two given measures are "close" to each other. In the IFS theory to determine a distance between measures usually the *Hutchinson metric* is used, which is defined by

$$d_H(\mu, \nu) = \sup_{f \in \text{Lip}(\leq 1)} \left\{ \left| \int_K f d\mu - \int_K f d\nu \right| \right\}, \quad (7)$$

where  $\text{Lip}(\leq 1)$  denotes the set of Lipschitz functions  $f: K \rightarrow \mathbb{R}$  with the Lipschitz constants not greater than 1 (with respect to the metric on  $K$ ). Assuming  $K$  to be compact, one can prove that  $d_H$  is indeed a metric on the compact space  $\mathcal{P}(K)$  of normalized Borel measures on  $K$ . Another noteworthy property of the Hutchinson metric is that the metric generates on  $\mathcal{P}(K)$  the same topology as the one induced by weak convergence of measures. One of the important consequences is that the

convergence of a sequence of measures in the sense of the Hutchinson metric implies the weak convergence of the sequence.

### 3 Lattice approximation of RIFS invariant measure

#### 3.1 Lattice-measure matching

So long as the pure theoretical lattice based analysis of a measure takes place, one does not have to take under consideration the problems related to the localization, orientation and extent of the support of the measure with respect to a coordinate system, because such an analysis utilizes  $\delta$ -lattice coverings of the whole space with arbitrarily small  $\delta$ . However, when there is a need to investigate a measure numerically, then, due to finite computer memory resources, we have only representations of *bounded* lattices with a restricted range of  $\delta$  at our disposal. Therefore, in practice, it is essential to match such a non-perfect representation of  $\delta$ -lattice to the support of the measure, so that the lattice is "filled" with the measure as much as possible. On the other hand, assuming the overlapping box approach to be used, Proposition 1 states that " $q$ -moment" properties of a measure are invariant under bi-Lipschitz transformations. Consequently, it is usually more convenient to transform the measure itself to fit in a given lattice of a bounded extent. Using (5) and (6) it is easy to check that, given the measure  $\mu$  specified by a RIFS  $(K; \{w_i\}_{i=1}^N; P)$ , the image of  $\mu$  under a bi-Lipschitz (and, hence, invertible and continuous) map  $f: X \rightarrow X$ ,  $K \subset X$ , is specified by the RIFS  $(f(K); \{f \circ w_i \circ f^{-1}\}_{i=1}^N; P)$ . (Moreover, by (4) it is readily seen that  $\text{spt } f^\# \mu = f(A)$  as expected.)

In practice, given a RIFS, it is convenient to transform the invariant measure by an invertible affine map (and, thus, bi-Lipschitzian) in order to obtain the image of the measure with the support bounded by the box of the unit side. Since the transformation of a measure implies the transformation of its support, the problem of the determination of an appropriate affine map is, in essence, equivalent to the determination of  $d$ -dimensional parallelepiped which bounds the RIFS attractor as tight as possible. (The problem is of the computational geometry character and as such is beyond the scope of this paper—for a heuristic algorithm see e.g.<sup>17</sup>.) Therefore, in the sequel we will assume the measure to be contained in the unit box.

#### 3.2 Approximation

Let  $(K; \{w_i\}_{i=1}^N; P)$ , where  $w_i: K \rightarrow K$  on compact  $K \subset \mathbb{R}^d$ , be an affine RIFS with the invariant measure  $\mu$  supported by the attractor  $A$ , such that  $A$  is a subset of the unit box  $C = \prod_{i=1}^d [0, 1]$ . Moreover, we assume that  $\mathbb{R}^d$  is endowed with the Euclidean norm, so the RIFS is contractive with respect to the Euclidean metric. (Note that a bi-Lipschitz transformation of a RIFS being contractive with respect to the Euclidean metric, we did in the previous section, may result in a RIFS which is *not* contractive w.r.t. the metric. Nevertheless, it is easy to show that there is still a metric induced by a norm on  $\mathbb{R}^d$  in which the maps of the transformed RIFS are contractions. Consequently, we can do the above assumption without any loss of generality, because all norms on  $\mathbb{R}^d$  are equivalent, i.e. they generate the same



topology.)

Our method of the approximation of RIFS invariant measures utilizes a recurrent recipe for expanding the measure into a sum of component measures supported by sets of diameters less than given  $\delta > 0$ . We used the idea of the decomposition itself previously in <sup>14</sup> to visualize 3D RIFS measures but in this paper we adapt it to the lattice context. In turn, the measure decomposition approach can be viewed as a RIFS generalization of the adaptive-cut algorithm for "usual" IFS invariant measures developed in <sup>18</sup>.

Let  $\mu = \sum_{i=1}^N \mu_i$  be the RIFS invariant measure (6). Since  $\text{spt } \mu \subset C$ , by the definition of the corresponding Markov operator (5) it is easy to show<sup>14</sup> that  $\mu$  can be expanded into a sum of "nonzero" measures

$$\mu_{i_1 \dots i_n} \equiv \left( \prod_{k=1}^{n-1} p_{i_{k+1} i_k} \right) (w_{i_1} \circ \dots \circ w_{i_{n-1}})^{\#} \mu_{i_n} \quad (8a)$$

$$\text{with } \text{spt } \mu_{i_1 \dots i_n} \subset w_{i_1} \circ \dots \circ w_{i_n}(C) \text{ and } \mu_{i_1 \dots i_n}(C) = \left( \prod_{k=1}^{n-1} p_{i_{k+1} i_k} \right) \pi_{i_n}, \quad (8b)$$

where each component measure  $\mu_{i_1 \dots i_n}$  can be further decomposed into the sum of  $m$  ( $m \leq N$ ) "nonzero" measures  $\mu_{i_1 \dots i_n i_{n+1}}$  such that  $p_{i_{n+1} i_n} > 0$ ,  $i_{n+1} \in \{1, \dots, N\}$ .

As a result of the recurrent use of the recipe above,  $\mu$  can be decomposed into a sum of measures supported by subsets of diameters less than a given  $\delta > 0$  with respect to the Euclidean metric. Unfortunately, the Euclidean metric is "unpleasant" for computing the diameters of sets (8b), because it requires, among others, the calculation of maximal eigenvalues of symmetric matrices in order to determine spectral norms of matrices specifying linear parts of the affine map compositions in (8). However all norms on  $\mathbb{R}^d$  are equivalent, so we can do such a decomposition of  $\mu$  with respect to any metric induced by a norm, even if the RIFS is not contractive in the metric chosen. By its "friendly" computing properties the maximum metric seems to be the best choice in the lattice context<sup>b</sup>. Since  $\text{spt } \mu$  is a subset of the unit cube  $C$ , by (8b) we obtain

$$\text{diam}(\text{spt } \mu_{i_1 \dots i_n}) \leq \|M_{i_1 \dots i_n}\|_{\infty} = \max_{i=1, \dots, d} \sum_{k=1}^d |m_{ik}|, \quad (9)$$

where  $M_{i_1 \dots i_n} = [m_{ik}]$  is the  $d \times d$  matrix specifying the linear part of  $w_{i_1} \circ \dots \circ w_{i_n}$ . As a result, taking advantage of the maximum metric we can obtain the decomposition of  $\mu$  into a sum of measures supported by subsets of the horizontal and vertical extents less than  $\delta$ .

Let  $L_n$  be the lattice covering of  $\mathbb{R}^d$  by  $d$ -dimensional boxes  $B = \prod_{i=1}^d [2^{-n} a_i, 2^{-n}(a_i + 1))$  of side  $2^{-n}$ . Let  $C_n$  be the subcollection of  $L_n$  specified by the lattice coordinates  $a_i \in \{0, \dots, 2^n - 1\}$ , that is, ignoring the "right-hand" boundary points of  $\prod_{i=1}^d [0, 1]$ ,  $C_n$  constitutes the  $2^{-n}$ -lattice covering of the unit

<sup>b</sup>It should be noted that in <sup>14</sup>, no lattices were employed, and to avoid the problem of computing diameters w.r.t. the Euclidean metric, the Frobenius norm was used.

cube  $C$ . Further, denote by  $\Omega_n$  the family of finite sequences  $i_1 \dots i_k$  (not necessarily of equal length) of indices  $i_j \in \{1, \dots, N\}$ , so that  $\mu = \sum_{i_1 \dots i_k \in \Omega_n} \mu_{i_1 \dots i_k}$  and  $\|M_{i_1 \dots i_k}\|_\infty < 2^{-n} \leq \|M_{i_1 \dots i_{k-1}}\|_\infty$ , where  $\mu_{i_1 \dots i_k}$  and  $\|M_{i_1 \dots i_k}\|_\infty$  are given by (8a) and (9) respectively. Now, we can define the function  $\phi: \Omega_n \rightarrow C_n$  by

$$\phi(i_1 \dots i_k) = \begin{cases} B : x_{i_1 \dots i_k} \in B & \text{if } x_{i_1 \dots i_k} \in \text{Int} C, \\ B : x_{i_1 \dots i_k} \in \overline{B} & \text{otherwise,} \end{cases} \quad (10)$$

where  $x_{i_1 \dots i_k}$  denotes the image of a point of the RIFS attractor subset  $A^{(i_k)}$  under the map composition  $w_{i_1} \circ \dots \circ w_{i_{k-1}}$  (if  $k = 1$ , we just take the identity map). It is easy to see that the definition is correct, because  $x_{i_1 \dots i_k} \in A \subset C$ , for by (4)  $w_{i_1} \circ \dots \circ w_{i_{k-1}}$  takes  $A^{(i_k)}$  to a subset of  $A^{(i_1)}$ , and the second case in (10) solves, in the unique way, the  $C$  vs.  $C_n$  boundary problem (so in both cases  $\phi$  chooses uniquely  $B$  from  $C_n$ ). Equipped with the above definitions, we define the  $2^{-n}$ -lattice approximation of  $\mu$  by the measure

$$\mathcal{U}^n = \sum_{i_1 \dots i_k \in \Omega_n} \mathcal{U}_{i_1 \dots i_k}, \quad (11a)$$

where the components  $\mathcal{U}_{i_1 \dots i_k}$  are measures that are uniform, lattice analogues of  $\mu_{i_1 \dots i_k}$  in (8), which are defined by

$$\mathcal{U}_{i_1 \dots i_k}(E) = \mu_{i_1 \dots i_k}(C) \cdot 2^{nd} \int_E \mathbf{1}_{\phi(i_1 \dots i_k)} dx \quad (11b)$$

whenever  $E \subset \mathbb{R}^d$  is Borel. Clearly,

$$\mathcal{U}_{i_1 \dots i_k}(C) = \mu_{i_1 \dots i_k}(C) \quad \text{and} \quad \text{spt} \mathcal{U}_{i_1 \dots i_k} = \overline{\phi(i_1 \dots i_k)}. \quad (11c)$$

**Proposition 2.** a)  $\mathcal{U}_{i_1 \dots i_k}(\phi(i_1 \dots i_k)) = \mu_{i_1 \dots i_k}(C) = \mu_{i_1 \dots i_k}((\phi(i_1 \dots i_k))_1)$ ,

b)  $|\int_K f d\mu_{i_1 \dots i_k} - \int_K f d\mathcal{U}_{i_1 \dots i_k}| \leq \sqrt{3d \cdot 2^{-n}} \cdot \mu_{i_1 \dots i_k}(C)$ , whenever  $f \in \text{Lip}(\leq 1)$  and  $K \subset \mathbb{R}^d$  is compact.

*Proof.* a) By (10) and (11b)  $\mathcal{U}_{i_1 \dots i_k}(\phi(i_1 \dots i_k)) = \mathcal{U}_{i_1 \dots i_k}(\overline{\phi(i_1 \dots i_k)})$ , so the first equality follows immediately from (11c). Now, we prove the second one. Denote  $C_{i_1 \dots i_k} := w_{i_1} \circ \dots \circ w_{i_k}(C)$  to shorten notation. We have  $x_{i_1 \dots i_k} \in C_{i_1 \dots i_k}$  and by (10)  $x_{i_1 \dots i_k} \in \overline{\phi(i_1 \dots i_k)}$ . Moreover,  $\text{diam}(C_{i_1 \dots i_k}) < 2^{-n}$  with respect to the maximum metric, because  $\|M_{i_1 \dots i_k}\|_\infty < 2^{-n}$  and  $\text{diam}(C) = 1$ . Hence, by the triangle inequality,  $\text{diam}(C_{i_1 \dots i_k} \cup \overline{\phi(i_1 \dots i_k)}) < 2 \cdot 2^{-n}$  with respect to the maximum metric. Thereby,  $C_{i_1 \dots i_k} \subset (\phi(i_1 \dots i_k))_1$ . But by (8b)  $\text{spt} \mu_{i_1 \dots i_k} \subset C_{i_1 \dots i_k}$ . Thus, the second equality holds.

b) First of all,  $f$  as a Lipschitz function is a Borel function and, thus,  $\mu_{i_1 \dots i_k}$  and  $\mathcal{U}_{i_1 \dots i_k}$  are measurable<sup>c</sup>. Further, by (11c),  $\text{spt} \mathcal{U}_{i_1 \dots i_k} \subset$

<sup>c</sup>It is the well-known fact that if  $w$  is a Borel map (i.e.,  $f^{-1}(E)$  is a Borel set whenever  $E$  is open) and  $\mu$  is a Borel measure, then  $w^\# \mu$  is also Borel. As  $w_i$  are Lipschitz continuous, they are Borel maps, and thus all  $\mu_{i_1 \dots i_k}$  are Borel measures, because  $\mu$  is Borel. The same is true for  $\mathcal{U}_{i_1 \dots i_k}$ , which being (rescaled) restrictions of the Lebesgue measure are naturally Borel.

$(\phi(i_1 \dots i_k))_1$ , so  $\int_K f d\mu_{i_1 \dots i_k} = \int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k}$ . Similarly,  $\int_K f d\mu_{i_1 \dots i_k} = \int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k}$ , because  $\text{spt } \mu_{i_1 \dots i_k} \subset (\phi(i_1 \dots i_k))_1$ , by the point (a) of this proposition. Now,  $(\phi(i_1 \dots i_k))_1$  is bounded and  $f \in \text{Lip}(\leq 1)$ , so  $f((\phi(i_1 \dots i_k))_1)$  is a subset of an interval  $[a, b]$  such that  $|b - a| \leq \text{diam}((\phi(i_1 \dots i_k))_1) = \sqrt{3d} \cdot 2^{-n}$  (w.r.t. the Euclidean metric). Hence, evidently  $\int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k} \in [a\mu_{i_1 \dots i_k}((\phi(i_1 \dots i_k))_1), b\mu_{i_1 \dots i_k}((\phi(i_1 \dots i_k))_1)]$  and analogously for  $\int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k}$ . As a result  $|\int_K f d\mu_{i_1 \dots i_k} - \int_K f d\mu_{i_1 \dots i_k}| = |\int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k} - \int_{(\phi(i_1 \dots i_k))_1} f d\mu_{i_1 \dots i_k}| \leq |b - a|\mu_{i_1 \dots i_k}(C)$  by the point (a) of this proposition.  $\square$

**Theorem 3.** Let  $\mu$  be a RIFS invariant measure with  $\text{spt } \mu \subset \prod_{i=1}^d [0, 1]$ , and let  $\mathcal{U}^n$  be the approximating measure defined by (11). Then the Hutchinson distance between  $\mu$  and  $\mathcal{U}^n$  obeys

$$d_H(\mu, \mathcal{U}^n) \leq \sqrt{3d} \cdot 2^{-n}.$$

*Proof.*  $d_H(\mu, \mathcal{U}^n) =$

$$\begin{aligned} & \sup\left\{\left|\int_K f d\mu - \int_K f d\mathcal{U}^n\right|\right\} = \sup\left\{\left|\int_K f d\left(\sum_{\omega \in \Omega_n} \mu_\omega\right) - \int_K f d\left(\sum_{\omega \in \Omega_n} \mathcal{U}_\omega\right)\right|\right\} = \\ & \sup\left\{\left|\sum_{\omega \in \Omega_n} \left(\int_K f d\mu_\omega - \int_K f d\mathcal{U}_\omega\right)\right|\right\} \leq \sup\left\{\sum_{\omega \in \Omega_n} \left|\int_K f d\mu_\omega - \int_K f d\mathcal{U}_\omega\right|\right\} \leq \\ & \sqrt{3d} \cdot 2^{-n}, \text{ where the suprema are over } \text{Lip}(\leq 1) \text{ functions, and the last inequality} \\ & \text{follows from Proposition 2(b), because } \sum_{\omega \in \Omega_n} \mu_\omega(C) = \mu(C) = 1. \end{aligned} \quad \square$$

On the basis of the above the sequence  $\{\mathcal{U}^n\}_{n=0}^\infty$  converges to  $\mu$  with respect to the Hutchinson metric. Let us note, however, that the distance between a measure and its approximation when expressed *only* in terms of the Hutchinson metric cannot be considered—as some authors seem to suggest—an objective measure of the quality of approximation in the space covering (e.g. lattice) context, because it does not take into account the size of the support of the measure relatively to the size of a covering element. (To see this, just scale a measure down to fit in a single element.) It seems that, both from the theoretical and numerical viewpoint, a good solution to this problem is to multiply  $d_H(\mu, \mathcal{U}^n)$  by  $\frac{n \log 2}{\log \#\{B \in C_n : \mathcal{U}^n(B) > 0\}}$ , because the value of the factor can be interpreted as a quantity of “stuffing”  $C_n$  by the support, and in the limit it is just the reciprocal of the generalized dimension  $D(0)$ —the box-counting dimension of the support.

#### 4 Computing Rényi dimensions

Let  $\mu$  be a RIFS invariant measure on  $K \subset \mathbb{R}^d$  with  $\text{spt } \mu \subset \prod_{i=1}^d [0, 1]$ , and let  $\mathcal{U}^n$  be the lattice  $L_n$  approximation of  $\mu$ , constructed in Sec. 3.2. Let  $G_\delta^\mu$  be the “standard” collection of boxes from the  $\delta$ -lattice  $L_\delta$  with positive  $\mu$  measure, i.e.  $G_\delta^\mu := \{B \in L_\delta : \mu(B) > 0\}$ . Further, define the  $\mathcal{U}^n$  analogue of  $G_\delta^\mu$  by  $G_n^{\mathcal{U}} := \{B \in L_n : \mathcal{U}^n(B) > 0\}$ . (Using the function  $\phi: \Omega_n \rightarrow C_n$  (10), it is readily seen that  $G_n^{\mathcal{U}}$  is just the image of the family  $\Omega_n$  of sequences defined in Sec. 3.2 under  $\phi$ , i.e.  $G_n^{\mathcal{U}} = \phi(\Omega_n)$ , and hence  $G_n^{\mathcal{U}} \subset C_n$ .) Since the convergence of measures

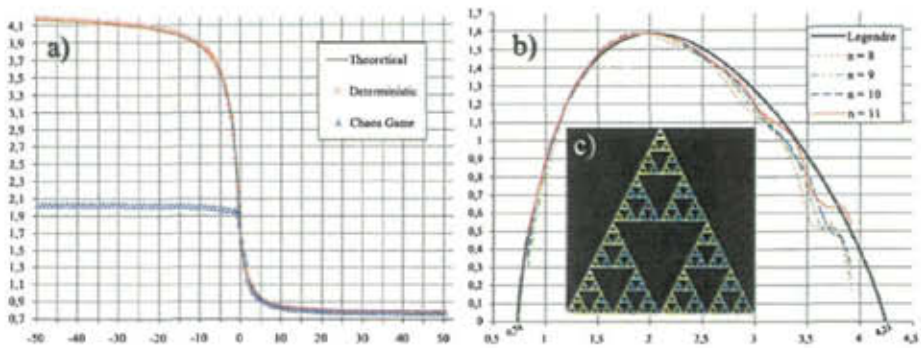


Figure 1. A RIFS measure supported by the Sierpiński Triangle: a) a comparison of estimations of generalized dimensions obtained with our method and the chaos game approach—the solid line represents the analytical result, b) the theoretical Legendre spectrum and coarse grained histograms, c) a draft of coarse grained Hölder exponents—the values of exponents are mapped to Hue component of HSV color space.

on a compact space with respect to the Hutchinson metric implies weak convergence, by Theorem 3 (and the fact that both real and imaginary parts of  $e^{itx}$  are continuous and bounded functions for all  $t \in \mathbb{R}$ ) we immediately obtain that the characteristic functions  $\mathbb{E}[\exp(iqU_n)]$ ,  $q \in \mathbb{R}$ , (and, thus, the Laplace transforms) of  $U_n = -\log \mathcal{U}_n(B)$ ,  $B \in G_n^\mu$ , converge (pointwise) to the characteristic function (respectively the Laplace transform) generated by  $\mu$ . Below we show that the use of the overlapping box approach allows one to state the same in the case of rescaled moment generating functions. We will need the following lemma:

**Lemma 4.** *Let  $\mu_\omega(B) > 0$ , where  $\omega \in \Omega_n$ ,  $B \in L_n$ , and  $\mu_\omega$  is one of the measures of the decomposition of  $\mu$  with respect to  $\Omega_n$ . Then  $(\phi(\omega))_1 \supset B$ , and, thus,  $\mathcal{U}_\omega((B)_1) = \mu_\omega(C)$ .*

*Proof.* By the construction of  $\mathcal{U}^n$  and Prop. 2(a), there is a box  $\phi(\omega) \in G_n^\mu$  such that  $\mu_\omega((\phi(\omega))_1) = \mu_\omega(C)$ , and it follows that  $\mu_\omega(K \setminus (\phi(\omega))_1) = 0$ . Because  $\mu_\omega(B) > 0$ , the last equality implies  $B \not\subseteq K \setminus (\phi(\omega))_1$ , so  $B \cap (\phi(\omega))_1 \neq \emptyset$ . But  $L_n$  is a lattice of disjoint boxes, therefore the nonempty intersection implies  $(\phi(\omega))_1 \supset B$ .  $\square$

**Theorem 5 (On approximation of  $\tau$ ).** *For all  $q \in \mathbb{R}$*

$$\limsup_{\delta \downarrow 0} \frac{\log \sum_{B \in G_\delta^\mu} \mu((B)_{\kappa_1})^q}{-\log \delta} = \limsup_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{B \in G_n^\mu} \mathcal{U}^n((B)_{\kappa_2})^q,$$

for any  $\kappa_1, \kappa_2 \in \mathbb{N}$ .

*Proof.* Denote by  $G_n^\mu$  the  $L_n$  lattice version of  $G_\delta^\mu$ , that is  $G_n^\mu := \{B \in L_n : \mu(B) > 0\}$ . Furthermore, to shorten notation, let  $S_q^\mu(n, \kappa) := \sum_{B \in G_n^\mu} \mu((B)_\kappa)^q$  and, similarly for  $\mathcal{U}^n$ ,  $S_q^\mu(n, \kappa) := \sum_{B \in G_n^\mu} \mathcal{U}^n((B)_\kappa)^q$ .

Let  $B \in G_n^\mu$ . So, by definition,  $\mu(B) > 0$  and, hence,  $\mu((B)_1) > 0$ . Decomposing  $\mu$  with respect to  $\Omega_n$ , we obtain  $\mu((B)_1) = \sum_{\omega \in \Omega_n} \mu_\omega((B)_1) > 0$ , so there are some component measures with  $\mu_\omega((B)_1) > 0$ . Since  $(B)_1$  consists of  $3^d$  boxes of the  $2^{-n}$ -lattice  $L_n$  and, naturally,  $\mu_\omega(C) \geq \mu_\omega((B)_1)$ , by Lemma 4 we obtain that  $\mathcal{U}_\omega((B)_2) \geq \mu_\omega((B)_1)$ , and thus  $\mathcal{U}^n((B)_2) \geq \mu((B)_1)$ . But  $B$  not necessarily has to be a member of  $G_n^\mu$ . However, by the assumption,  $B \in G_n^\mu$ , so there is a component measure  $\mu_\omega$  of the decomposition of  $\mu$  with  $\mu_\omega(B) > 0$ , and it follows (again by Lemma 4) that there is  $\phi(\omega) \in G_n^\mu$  such that  $(\phi(\omega))_1 \supset B$ . Thereby  $(\phi(\omega))_3 \supset (B)_2$  and, as a result, for each  $B \in G_n^\mu$ , there is  $\phi(\omega) \in G_n^\mu$  such that  $\mathcal{U}^n((\phi(\omega))_3) \geq \mu((B)_1)$ . The correspondence, however, does not have to be one-to-one. Nevertheless, the number of boxes from  $G_n^\mu$ , which the same box  $\phi(\omega)$  can be assigned to, is bounded by the number of boxes from  $L_n$  that  $(\phi(\omega))_1$  consists of, namely  $3^d$ . Consequently, repeating each term  $3^d$  times in  $S_q^\mu$  and  $S_q^\mu$  respectively yields the inequalities  $S_q^\mu(n, 1) \leq 3^d S_q^\mu(n, 3)$  for  $q \geq 0$  and, respectively,  $3^d S_q^\mu(n, 1) \geq S_q^\mu(n, 3)$  for  $q < 0$ . As a result, we get

$$\limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 1)}{-\log 2^{-n}} \leq \limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 3)}{-\log 2^{-n}} \quad \text{for } q \geq 0, \quad (12a)$$

and the opposite inequality for  $q < 0$ .

Now, let  $B \in G_n^\mu$ . Then, by the definition of  $\mathcal{U}^n$ ,  $\mathcal{U}^n((B)_1)$  is equal to the sum of  $\mathcal{U}_\omega((B)_1)$ , where  $\omega \in \phi^{-1}((B)_1)$ . Next, using Prop. 2(a), we show that for each  $\mathcal{U}_\omega$ ,  $\mathcal{U}_\omega((B)_1) \leq \mu_\omega((B)_2)$  and, hence,  $\mathcal{U}^n((B)_1) \leq \mu((B)_2)$ . Then, the analogous reasoning to that used in the previous case, leads to the inequalities

$$\limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 1)}{-\log 2^{-n}} \leq \limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 3)}{-\log 2^{-n}} \quad \text{for } q \geq 0, \quad (12b)$$

and the opposite one for  $q < 0$ .

Since  $\delta_n = 2^{-n}$  satisfies  $\delta_n \geq \delta_{n+1} \geq c\delta_n$  for  $c \geq \frac{1}{2}$ , using Prop. 1 we obtain  $\limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 3)}{-\log 2^{-n}} = \limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 1)}{-\log 2^{-n}} = \limsup_{\delta \downarrow 0} \frac{\log S_q^\mu(n, \kappa_1)}{-\log \delta}$ , and  $\limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 3)}{-\log 2^{-n}} = \limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, 1)}{-\log 2^{-n}} = \limsup_{n \rightarrow \infty} \frac{\log S_q^\mu(n, \kappa_2)}{-\log 2^{-n}}$ ,  $\kappa_1, \kappa_2 \in \mathbb{N}$ . Hence, by (12) we get the thesis.  $\square$

## 5 Discussion of results

In order to test the accuracy of our method, as the first example we choose the RIFS measure specified by  $P = \begin{bmatrix} 1/16 & 5/16 & 5/8 \\ 6/13 & 3/13 & 4/13 \\ 4/19 & 14/19 & 1/19 \end{bmatrix}$  and supported by a Sierpiński Triangle. Since the RIFS satisfies the open set condition, it can be regarded as a special case of the Mauldin-Williams fractal construction and, thus, the generalized dimensions can be determined analytically by means of the auxiliary functions theory<sup>4</sup>. Here,  $\tau(q)$  is defined as the unique solution of the equation  $\lambda_{\max}([0.5^{\tau(q)} p_{ik}^q]) = 1$ , where  $\lambda_{\max}(\cdot)$  is the spectral radius of a matrix. Fig. 1(a) depicts the comparison of estimations of generalized dimensions obtained with our method and the chaos

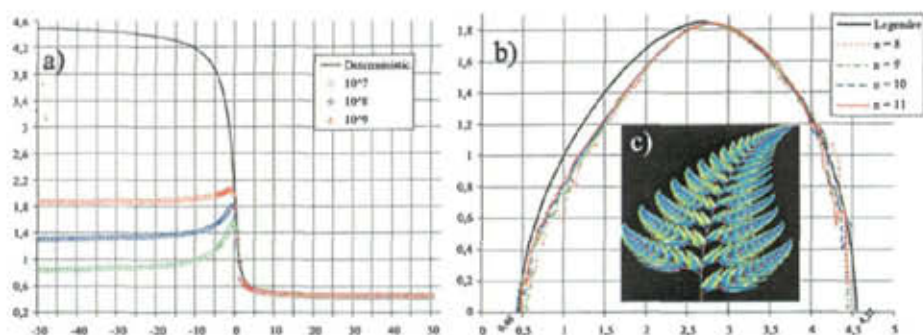


Figure 2. A RIFS measure supported by the Barnsley fern: a) generalized dimensions—a comparison of outcomes supplied by our method and the chaos game approach for different numbers of steps, b) the theoretical Legendre spectrum and coarse grained histograms, c) a draft of coarse grained Hölder exponents.

game approach, respectively. The lattice of  $2^{11} \times 2^{11}$  boxes was used, and to approximate the measure with the chaos game, the huge number of  $10^8$  iterations were performed. While both approaches have provided good results for  $q \geq 0$ , however, as one may have expected, due to relatively small diagonal entries of  $P$  the chaos game failed for negative  $q$ . Another noteworthy characteristic of this example is that  $\tau$ , as arising from the Mauldin-Williams construction, is an *analytic* function of  $q$  and, hence, differentiable. Consequently, the Gärtner-Ellis theorem<sup>15</sup> applies and, thus, the Legendre and large deviation spectra are equal, i.e. the multifractal formalism holds. The convergence of the coarse grained histograms (determined from the lattice approximation of the measure supplied by our method) to the Legendre spectrum seems to be convincing, Fig. 1(b), but the "bumps" for  $\alpha > 2$  cannot be unnoticed (probably affected by the relatively large range of values in  $P$ ). Finally, the asymmetry of the spectrum reflects the "recurrent" character of the RIFS measure, because the "usual" Sierpiński-Triangle IFSs with probabilities specify measures with symmetrical spectra.

The second example is the classical IFS of the Barnsley fern<sup>10</sup>, which we equipped with the probability matrix  $P = \begin{bmatrix} 1/7 & 3/14 & 1/2 & 1/7 \\ 4/7 & 1/7 & 3/14 & 1/14 \\ 1/15 & 8/15 & 1/5 & 1/5 \\ 1/5 & 1/15 & 2/15 & 9/15 \end{bmatrix}$  to get a RIFS measure. Since the IFS is affine and, in addition, overlapping (see Fig. 2(c) for a draft of the measure after a transformation to fit in the square), there are no formulas to determine the generalized dimensions analytically in this case. Estimations of generalized dimensions of the measure obtained from our method and the chaos game are depicted in Fig. 2(a). The good results from the previous example as well as the noticeable but slow convergence of the chaos game outcomes towards the area occupied by those supplied by our method allows us to assert that again our approach has turned out much more accurate (and much more efficient if one takes into account the enormous numbers  $10^7$ ,  $10^8$ , and  $10^9$  of the chaos game steps necessary to produce the presented results). Just like previously, the lattice



of  $2^{11} \times 2^{11}$  boxes was used. Fig. 2(b) shows the Legendre spectrum and coarse grained histograms computed on the basis of our method of approximating a RIFS measure.

In the two examples the overlapping box approach ( $\kappa = 1$ ) was used both for counting boxes filled with our approximation of a measure and those by the chaos game respectively (in the latter case to overcome the negative  $q$  problems mentioned in Sec. 2.1). Also in both examples the final results were obtained by means of linear regression over  $n = 5, \dots, 11$ . Finally, due to accuracy of our method, the measures of some boxes can be extremely small numbers (especially in the case of high-resolution lattices), so as to avoid numerical range problems and improve numerical behavior for large negative  $q$ , we additionally scaled up the values in the lattice and then used the formulas modified adequately to take into account the scaling factor applied (in spite of the fact that the scaling, being independent of  $n$ , has no impact on the limit, yet in practice we deal only with a finite range of  $n$ ).

### Acknowledgments

My thanks go to Prof. Jan Zabrodzki at Computer Graphics Laboratory of Warsaw University of Technology for reading the original version of this paper and support.

### References

1. H. Hentschel and I. Procaccia, *Physica D* **8**, 435 (1983).
2. R.H. Riedi and B.B. Mandelbrot, *Adv. Appl. Math.* **16**, 132 (1995).
3. R. Cawley and R.D. Mauldin, *Adv. Math.* **92**, 196 (1992).
4. G.A. Edgar and R.D. Mauldin, *Proc. London Math. Soc.* **65**, 604 (1992).
5. K.J. Falconer, *J. Theor. Probab.* **7**, 681 (1994).
6. L. Olsen, *Random Geometrically Graph Directed Self-Similar Multifractals*, (Longman Scientific & Technical, Harlow, 1994).
7. M. Arbeiter and M. Patzschke, *Math. Nachr.* **181**, 5 (1996).
8. J. King, *Adv. Math.* **116**, 1 (1995).
9. L. Olsen, *Math. Proc. Cambridge Philos. Soc.* **120**, 709 (1996).
10. M.F. Barnsley, *Fractals Everywhere*, (Academic Press, New York, 1988).
11. R. Pastor-Satorras and R.H. Riedi, *Physica A* **29**, L391 (1996).
12. J. Elton, *Ergod. Th. and Dynam. Sys.* **7**, 481 (1987).
13. G. Froyland and K. Aihara, *J. Bif. and Chaos* **10**, 103 (2000).
14. T. Martyn, *A method for visualizing 3D Recurrent IFS invariant measures*, submitted 2003.
15. R.H. Riedi, *J. Math. Anal. Appl.* **189**, 462 (1995).
16. M.F. Barnsley, J.H. Elton and D.P. Hardin, *Constr. Approx.* **5**, 3 (1989).
17. T. Martyn, *Comp. & Graph.* **27**, 535 (2003).
18. D. Hepting, P. Prusinkiewicz, and D. Saupe, in *Proc. Fractal '90*, (IFIP, Lisbon, 1990).

# NONLINEAR DYNAMICS AND PREDICTION OF THE CASPIAN SEA LEVEL

N. G. MAKARENKO, L. M. KARIMOVA AND Y. B. KUANDYKOV

*Institute of Mathematic, 125 Pushkin street, Almaty,  
480100, Kazakhstan*

*E-mail: makarenko@math.kz*

M. M. NOVAK

*Kingston University, School of Mathematics, Kingston University,  
Surrey KT1 2EE, England*

*E-mail: novak@kingston.ac.uk*

The Caspian Sea is the largest intracontinental reservoir without water outflow and thus demonstrates the unique global evolution extending over a huge time interval. The economies of communities around the Caspian Sea are highly dependent on sea levels. Therefore, development of correct models for the sea level prediction is very important. Existing linear models of reservoirs without outflow and also non-linear stochastic models are based on hydrologic balance equation and have a number of limitations. We have applied the embedology technique to study the dynamics of the Caspian Sea level and have demonstrated bistability of the level dynamics for various time scales. Fractal approximation allows us to obtain an extended equidistant annual time series, which is used for global nonlinear Caspian Sea level forecasting with the help of the Artificial Neural Network.

## 1 Introduction

The Caspian Sea is the largest intracontinental reservoir without water outflow. On geological scales, the history of the sea is represented by the repeated change of transgressive and regressive phases<sup>1</sup>. That is reflected by faint traces of paleodata, by scanty historical information and also by the monitoring on the relatively short period of measurement. The regular monthly Caspian Sea level (CSL) data are available only since 1837<sup>a</sup>. The Caspian region is lowlands, thus even slight rise of the water level may result in floods covering huge areas near the sea. Sometimes big fast moving waves of water come to coasts with speed of a train, penetrating inland for up to 100 km. During the last Caspian Sea regression of 1883–1977, the level has dropped up to -26.4 m<sup>b</sup> and large sea area, of about 100 km<sup>2</sup>, became dry land<sup>2</sup>. Then the transgression started, from 1978, and by the beginning of 1996 the water level had risen by 2.4 m. The area of the sea had increased by 40,000 km<sup>2</sup>. At present, a stabilization is observed. It should be noted that, as researches<sup>4</sup> have shown, the regression observed over 1883–1977 as well as the transgression of 1978–1996 can not be reasonably explained by influencing on CSL of anthropogenic factors such as irrigation, draining etc.

The Caspian region is a huge oil region and the environment and the economies

<sup>a</sup>there are two variants of the instrumental CSL data: annual data beginning from 1830 and monthly data beginning from January 1837

<sup>b</sup>the sign "minus" means that the CSL is below the World Ocean level; more precisely, the CSL is given according to the so-called Baltic reference system<sup>3</sup>



of that region are highly dependent on the nearest episode of the sea level dynamics. In the event of further sea level rises, the territory of the coastal cities, villages with farm fields, roads and railways are likely to be flooded. Therefore, development of correct models of the sea level prediction is very important.

At the first sight, the problem of modelling the dynamics of such a reservoir, without outflow, is not so difficult<sup>3</sup>. But simple schemes of probabilistic prediction based on balance arithmetic of the flow and evaporation were not successful: the correlation connections of balance components are very simplified and the evaporation is not described by correct methods<sup>5</sup>. Consequently, an alternative approach to modelling and prediction of dynamic regimes of sea level, and based on the chaotic dynamics, is applied in this paper. The structure of this paper is as follows. In Section 2 we describe briefly the Caspian Sea level data used in our investigation. Moreover, their dynamical and fractal characteristics, the reconstruction of the topological model and fractal approximation of historical data are also to be found there. The results of CSL predictions are presented in Section 3. The summary is found in the concluding section.

## 2 Caspian Sea Level Data Description and Methods of Analysis

There are three types of the Caspian Sea level data, in particular instrumental<sup>7</sup>, historical<sup>6</sup> (600 B.C.-1760 A.D.) and ancient (starting from 14,000 B.C.) time series<sup>1</sup>. The last group of data is not used in our investigation. The Caspian Sea level data have been measured only since 1830 and they do not contain the information about global variations of the sea level (cf. Figure 1). The historical time series was obtained by indirect methods: historical information, <sup>14</sup>C dating and with the help of regression between CSL and temperature of the Volga basin, and between CSL<sup>c</sup> as well as the oceans' level<sup>6,1</sup>. These data are nonequidistant and contain essential errors of observation concerning absolute radiocarbon dating and inherent uncertainties of regression relationships. Nevertheless, the synthesized equidistant decennial time series was constructed and used by Russian researches<sup>1</sup>.

### 2.1 Reconstruction of the Topological Model

Contemporary investigation of the processes influencing CSL in a frame of the 'sea-atmosphere-land' system shows that the Caspian Sea represents a nonlinear system. It has a few meta-steady states with noise-induced transitions among them<sup>5</sup>. This is the reason, why application of nonlinear dynamics for data analysis and prediction is necessary.

The idea of the dynamics reconstruction from time series is based on the following statements.<sup>8,9,10</sup> Let  $M(x)$  is a differential manifold, and the map  $T : M \rightarrow M$  is a diffeomorphism, when the underlying dynamics is described by differential equations. It is said that  $M$  is a phase space of the smooth dynamical system  $T$ . The observation function  $h : M \rightarrow R$  converts a trajectory of the system  $T : (x_0, T(x_0), T^2(x_0), \dots) = (x_0, x_1, x_2, \dots)$  into a time series or sequence of records

<sup>c</sup>regression was used for CSL reconstruction starting from 1760

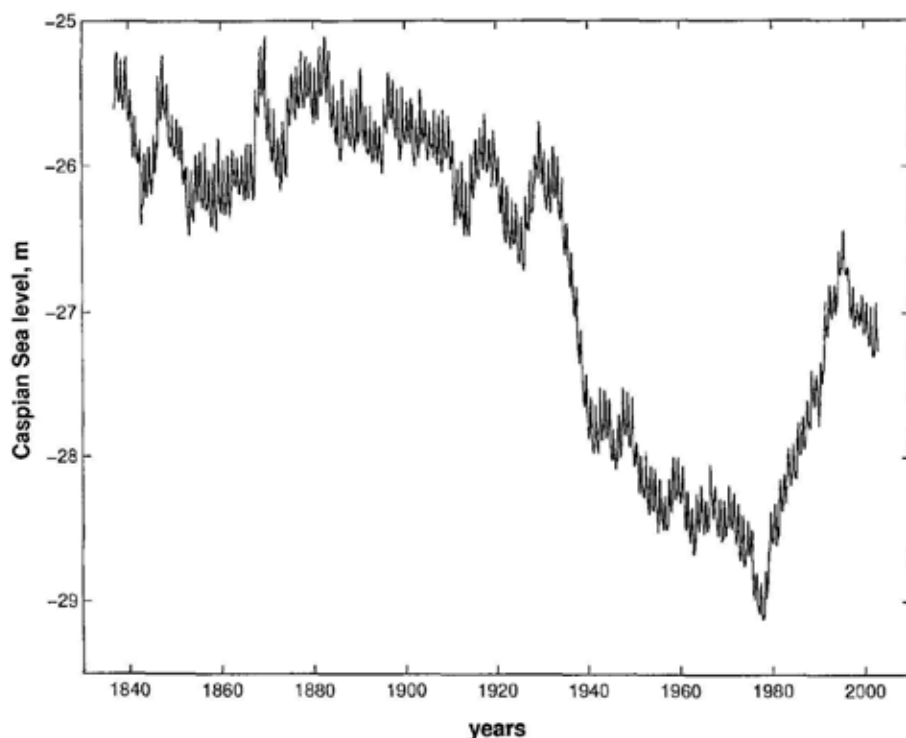


Figure 1. Mean monthly Caspian Sea level data from 1837 till 2002.

$h_0, h_1, h_2, \dots$ , where  $h_n = h(T^n(x_0)) \equiv h(x_n)$ . Let a trajectory  $T^n(x)$  after sufficiently large  $n$  belong to the  $d$ -dimensional attractor. Let us define the delay coordinates map  $\Phi(T, h) : M \rightarrow R^m$  as

$$\Phi(T, h)(x_k) = (h(x_k), h(T(x_k)), \dots, h(T^{m-1}(x_k))) \equiv \mathbf{h}_k. \quad (1)$$

Then, according to the Takens theorem<sup>11</sup>, it is a generic property that  $\Phi(T, h)$  is an embedding, when dimension  $m > 2d$ . A corresponding dynamical system  $F : \Phi(M) \rightarrow \Phi(M)$ , where

$$F(\mathbf{h}_k) = \Phi \circ T \circ \Phi^{-1}(\mathbf{h}_k) = \Phi \circ T \circ \Phi^{-1}(\Phi(x_k)) = \Phi \circ T(x_k) = \Phi(x_{k+1}) = \mathbf{h}_{k+1} \quad (2)$$

in delay space  $\Phi(M) \subset R^m$  is a diffeomorphic copy of the original system  $T : M \rightarrow M$ . Recall that a smooth map:  $\Phi : X_1 \rightarrow X_2$ , where  $X_1$  and  $X_2$  are smooth manifolds, embeds  $X_1$  in  $X_2$  (is an "embedding"), if  $\Phi$  is a diffeomorphism from  $X_1$  to a smooth submanifold of  $X_2$ , where  $X_2$  is called the embedding space with dimension  $\dim X_2 \neq \dim X_1$ . One can think of  $\Phi(X_1)$  as being the realization of  $X_1$  as a submanifold  $X_2$ , in other words the topological structures of  $X_1$  and  $\Phi(X_1) \subset X_2$  are diffeomorphically equivalent<sup>9</sup>. It means that if we can find the embedding  $\Phi : M \rightarrow R^{2d+1}$ , then we can analyze the structure of the dynamical system trajectory in  $R^{2d+1}$  and then readily deduce the properties of the actual trajectory

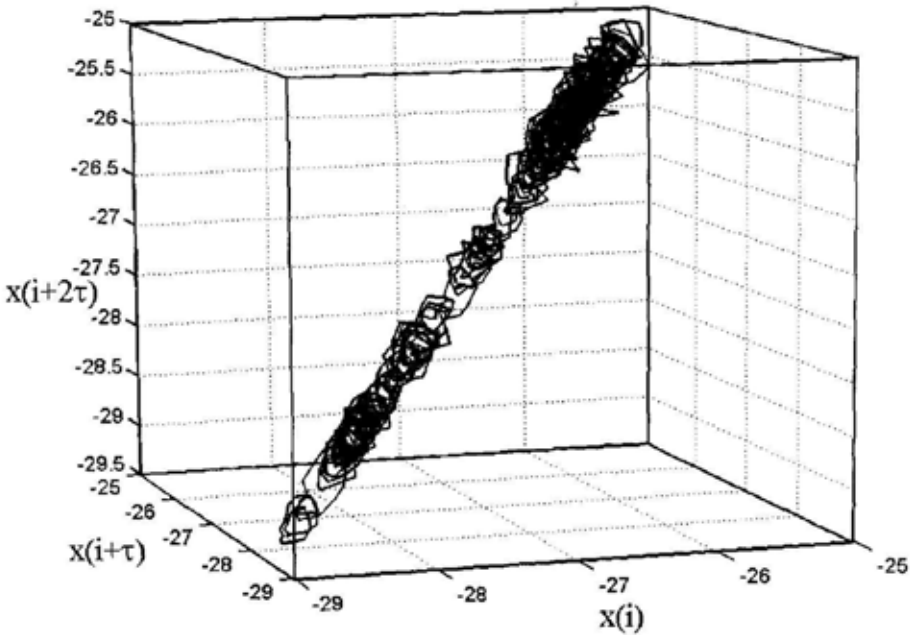


Figure 2. The phase portrait of monthly CSL time series.

on the attractor in  $M$ . So, in particular, Lyapunov exponents and entropy of system can be extracted from the reconstruction.<sup>12,13</sup> In practice, the components of vectors  $\mathbf{h}_k \in R^m$  must be independent, thus time series records are chosen with an appropriate value of the lag  $\tau$ :  $\mathbf{h}_k = (h_k, h_{k-\tau}, \dots, h_{k-(m-1)\tau})$ . The proper value  $m$  of the embedding dimension is estimated by means of well-known algorithms.<sup>12,14</sup>

For monthly instrumental CSL time series we have estimated that  $m \simeq 8 - 10$ . Figures 2 and 3 demonstrate the 3D-delay reconstructions of the instrumental and historical decennial time series (note that 3D reconstruction guarantees only immersion<sup>d</sup>, but not embedding). Both reconstructions indicate that the CSL dynamics is bistable as there are two essential states in the phase portraits connected with high and low values of the CSL.

## 2.2 Fractal Approximation of Historical Data

Historical decennial CSL time series (600B.C.-2000A.D.) contains about 260 records. However, this number of records is insufficient for a long-term CSL prediction, in particular for 10 years and more, a period that is of great practical interest. One way to avoid the data shortage is to use historical annual time series<sup>7</sup> obtained from decennially sampled historical time series by means of linear interpolation. However, in this case the CSL fluctuations on time intervals shorter than 10 years are totally lost. Thus, for proper reconstruction of annual CSL data

<sup>d</sup>i.e. differential of  $\Phi$  is one-to-one map

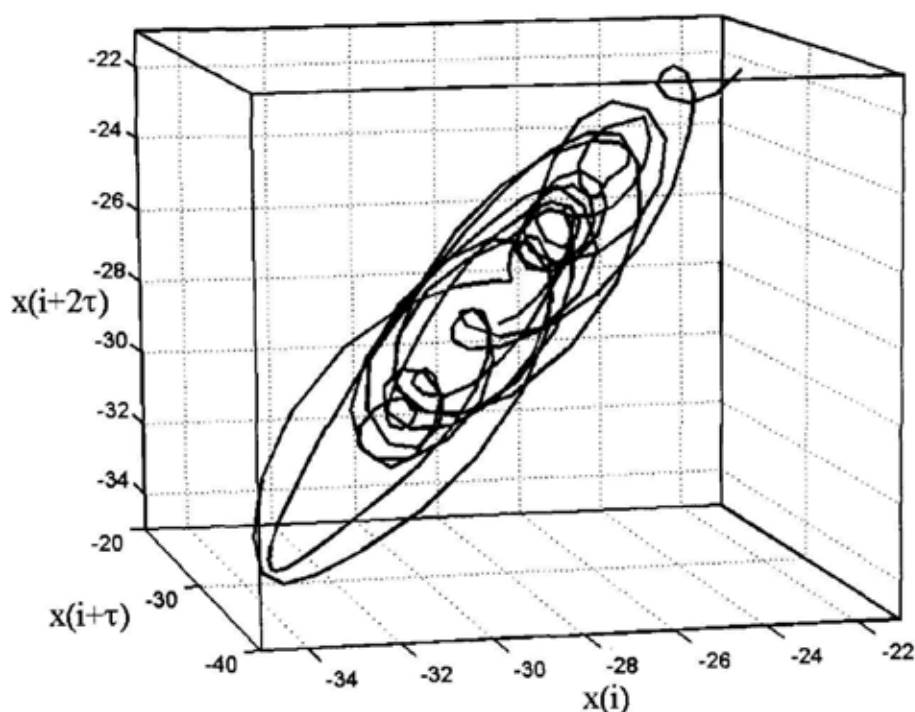


Figure 3. The phase portrait of decennial CSL time series.

we used the method of fractal approximation<sup>15</sup>. It is stated that using such a method is optimal, when

- there is no prior information on data structure corresponding to the short scales, and
- initial time series exhibits fractal properties<sup>e</sup>.

In our approach we rely upon the following facts. First, the Hurst rescaled range method points out that CSL has fractal structure as the Hurst exponent of CSL increments equals 0.767<sup>17</sup>. Second, bistability of the CSL dynamics on various time scales (cf. Figures 2, 3) indicates that there is a self-similarity in CSL behaviour. Third, the instrumental CSL time series has a well-defined multifractal spectrum. Figure 4 demonstrates both such multifractal spectra<sup>18</sup>: the Legendre spectrum and the large deviation spectrum, obtained by means of FracLab software<sup>19</sup>. Consequently, we assume that fractal scaling properties of CSL are invariant under transition from decennial scale to annual.

We pose the fractal interpolation problem<sup>15,16</sup> with a set of input points  $\{(x_i, y_i)\}_{i=0}^N$  with nodes  $0 = x_0 < x_1 < \dots < x_N = 1$  and ordinates  $y_i = F_{app}(x_i) \in R$  assuming some continuous function  $F_{app} : [0, 1] \rightarrow R$ . Classically,

<sup>e</sup>for example, variogramme behaves as  $E[(x(t_1) - x(t_2))^2] \propto (t_1 - t_2)^\alpha$

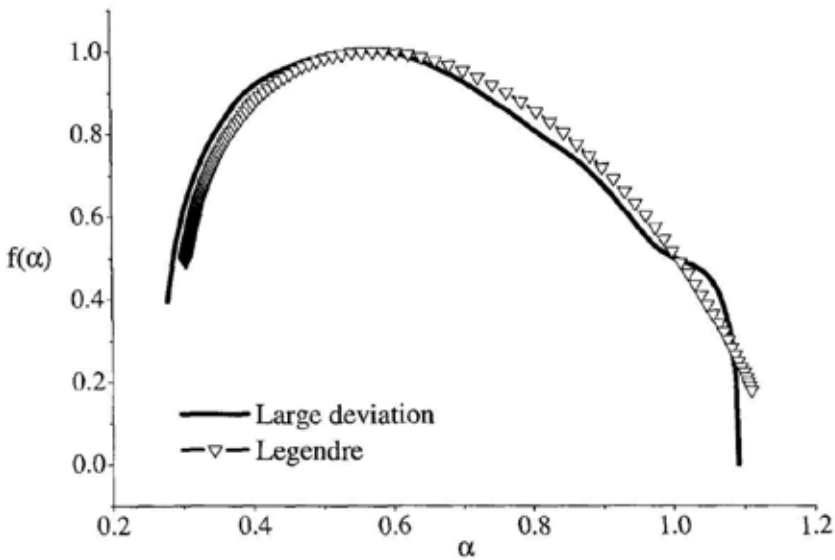


Figure 4. Multifractal spectra for instrumental CSL time series. Abscissa axis corresponds to singularity exponent  $\alpha$ , while ordinate axis to dimension  $f(\alpha)$  of fragments of measure with singularity  $\alpha$ .

if  $F_{app}$  is assumed smooth, then the input points are interpolated globally with a single polynomial of degree  $N$ , or piecewise with low-degree polynomials. An alternative assumption is that the interpolation function  $F_{app}$  is self-similar, and typically not smooth but fractal. Such a function is called a *Fractal Interpolating Function* (FIF) <sup>15</sup>. Let  $Gr(F_{app}) = \{(x, y) | y = F_{app}(x)\}$  stands for the graph of  $F_{app}$ .

We construct an *Iterated Function System* (IFS) whose attractor is the graph  $Gr(F_{app})$ . For  $i = 1, 2, \dots, N$ , let  $T_i : [0, 1] \times R \rightarrow [0, 1] \times R$  be affine transformations of the form <sup>16</sup>:

$$T_i : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} a_i & 0 \\ b_i & c_i \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d_i \\ e_i \end{bmatrix}, \quad (3)$$

where  $|c_i| < 1$  is given as a parameter controlling the roughness of the function, and  $a_i, b_i, d_i$  and  $e_i$  are determined by either the constraints

$$\begin{aligned} T_i(0, y_0) &= (x_{i-1}, y_{i-1}), \\ T_i(1, y_N) &= (x_i, y_i), \end{aligned} \quad (4)$$

or the “reflected” constraints

$$\begin{aligned} T_i(1, y_N) &= (x_{i-1}, y_{i-1}), \\ T_i(0, y_0) &= (x_i, y_i). \end{aligned} \quad (5)$$

Choosing the appropriate metrics it can be easily shown that each  $T_i$  is a contractive map in the corresponding metric space. Hence, by the fixed point theorem, there exists one and only one function  $F_{app}$  satisfying the invariance

$$Gr(F_{app}) = \bigcup_i T_i(Gr(F_{app})), \lim_{n \rightarrow \infty} \mathbf{T}^n(A) = Gr(F_{app}), \quad (6)$$

where  $\mathbf{T} = \bigcup T_i$ ,  $\mathbf{T}^n = \mathbf{T}(\mathbf{T}^{n-1})$ , and  $A$  is a compact set from  $R^2$ . Finally, we obtain the resulting graph  $Gr(F_{app})$  having a much better resolution in time.

Our approach was based on the fact that the reconstructions of historical decennial data, implemented by various researches, have considerable divergences for the 1540-1590 time period. In particular, according to researchers<sup>1</sup>, there was a considerable minimum of CSL at -31.0 m in 1560, however, such a value absents in CSL data given by another information source<sup>7</sup>, where CSL value in 1560 has been estimated around -27.1 m. Consequently, we implemented our own reconstruction of the fragment of 1540-1590 in the original historical time series<sup>1</sup>, relying upon nonlinear dynamics of the Caspian Sea. The data of this fragment were regarded as "missing" and were recovered with the help of the method of modelling the missing data in small-dimensional manifolds<sup>21</sup>. Fractal approximation were applied to the both, original and recovered, time series. Obtained data and their differences are shown in Figure 5.

Both variants of fractal approximation were chosen on condition that they were fitted to instrumental annual CSL data in the best way<sup>f</sup>. At the final stage we included records of the actual annual instrumental measurements from 1837 to 2002 years instead of modelled ones. So, we obtained two time series of 2603 records in length to predict future annual CSL data. To estimate the quality of the fractal approximation, the Hölder exponents of time series were computed<sup>20</sup>. We found that there was no marked distinction between the Hölder exponent behaviour of the instrumental time series and the fractal approximation, i.e the synthesized time series has uniform regularity. The Hölder pointwise exponent mean values of both obtained time series were evaluated as  $(0.77 \pm 0.20)$  and  $(0.84 \pm 0.24)$ , respectively.

### 3 CSL Forecasting by Artificial Neural Networks

Takens' algorithm enables a construction of a nonlinear predictor<sup>12,23</sup>:  $h_{k+\tau} = F(\mathbf{h}_k)$ . As regards function  $F$ , it is known only that this function is nonlinear, continuous and probably differentiable. For such predictor approximation the local and global approaches are used. So, Farmer *et al*<sup>22</sup> applied local linear techniques, where a local neighbourhood containing  $\mathbf{h}_k$  was used for prediction. The global nonlinear approximators  $\hat{F}(\mathbf{h}_k, \mathbf{w}) : R^m \rightarrow R$  can be defined as  $\hat{F}(\mathbf{h}_k, \mathbf{w}) = \sum w_i \phi(\|\mathbf{h}_k - \mathbf{h}_i\|)$ , where  $\phi$  is either radial basis functions, i.e. "gaussian",

<sup>f</sup> the genetics algorithm was used for this purpose

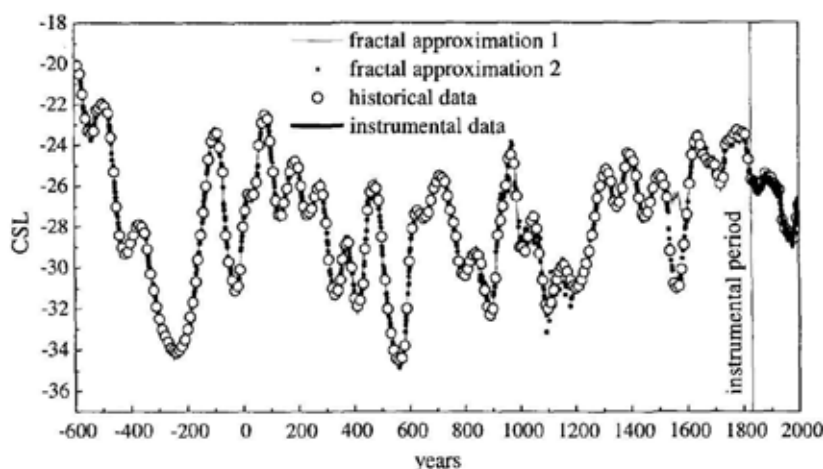


Figure 5. The historical CSL(m) data and annual obtained by fractal approximation together with instrumental ones.

or "sigmoid" functions, for example,  $\sigma(x) = \tanh wx$ . In the latter case, we obtain Neural Networks-approximator<sup>25</sup>, which has only one hidden layer in its simplest variant:  $\hat{F}(\mathbf{h}_k, \mathbf{w}) = \sigma(\sum_n w_n \sigma(\sum_j w_j h_j))$ . In general, the Artificial Neural Networks (ANN) are extremely flexible apparently capable of approximating very complicated multivariate functions with the help of superposition of standard (sigmoid) univariate functions<sup>24,25</sup>. In this work we used multilayer ANN with feed-forward training algorithm for prediction of CSL basing on historical and instrumental time series.

### 3.1 The Results of Monthly and Annual CSL data Predictions

The predictions were made on the basis of monthly instrumental data and annual time series obtained from the historical ones with the help of fractal approximation. The training set was constructed as an  $m$ -dimensional delay vectors table. The various values of reconstruction parameters were used. In particular, for the monthly instrumental time series they were  $m = 9; \tau = 80$  and  $m = 29; \tau = 84$  for the annual time series.

The example of prediction of the monthly CSL data is shown in Figure 6. The records from 1837 to the end of 1999 were used for the construction of the training set. All available data from December 1999 were used by the network for testing, i.e. those points were not included in the training set. The prediction was obtained for a whole time interval from January 2003 to August 2009, simultaneously with the test. Note that using  $\tau > 1$  for prediction model gives the possibility to obtain a vector prediction, i.e. the number of simultaneously predicted values equals to  $\tau$ .

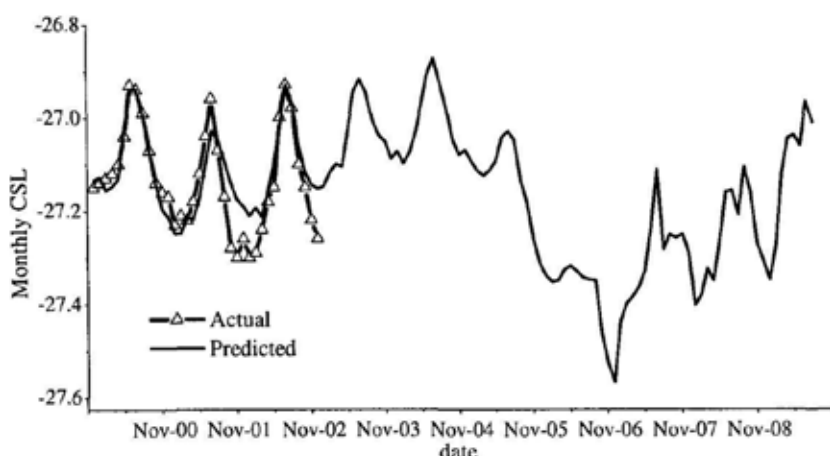


Figure 6. The example of monthly CSL(m) prediction.

The Lyapunov prediction horizon for this monthly time series equals 50 months, but this estimation is not much reliable as the concept of Lyapunov exponents is defined only for deterministic systems. For stochastic dynamical systems, however, a naive application of known algorithms<sup>26,27</sup> for the estimation of all or at least the largest Lyapunov exponent can give spurious positive exponents<sup>28</sup>.

The predictions of yearly Caspian Sea Level time series were made in the same way. The training sets were constructed from both variants of fractal approximations of the historical time series. Predictions were made with different values of a lag, so that the length of the predictions varied from 45 to 84 points. Figure 7 demonstrates the set of variants of predictions up to 40 years.

Presence of a number of predictions here can be explained by the following way. Let  $g = \{\{x_i, y_i\} \in R^d \times R\}_{i=1}^N$  is a set consisting of data, which has been obtained by random sampling a function  $f$ , belonging to some space of functions  $X$ . The task is to recover the function  $f$ , or an estimate of it, from data of the set  $g$ . But this problem is clearly ill-posed<sup>29</sup>, since it has an infinite number of solutions minimizing approximation error:  $\sum_i (f(x) - y_i)^2 \rightarrow \min$ . In the case of ANN, each solution (i.e. prediction) corresponds to individual initialization procedure of ANN weights. In order to choose one particular solution we need to have some a priori knowledge of the function that has to be reconstructed. For example, form of a priori knowledge often consists in assuming that the function is smooth, in the sense that two similar inputs correspond to two similar outputs. So, regularization theory maintains that the solution of the ill-posed problem can be obtained from a variational principle, which contains both the data and prior smoothness information. However, we did not use any regularization technique for predictions demonstrated here.



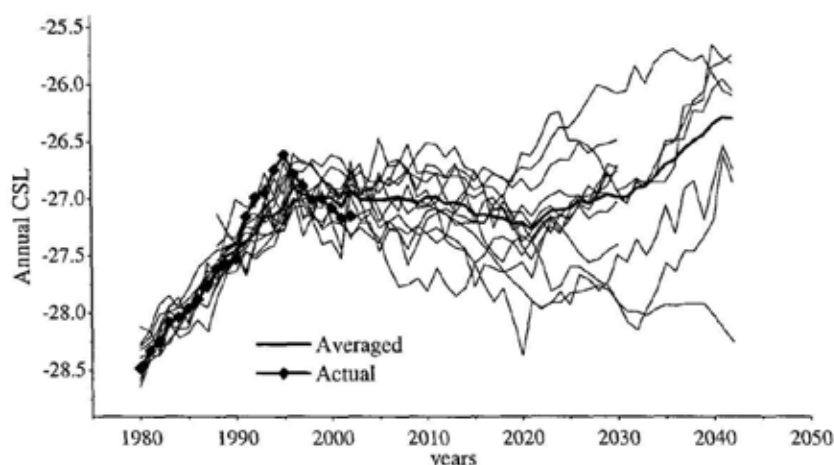


Figure 7. The set of predictions of the annual CSL(m) time series. The thick black curve was obtained through averaging of these predictions.

#### 4 Conclusions

Our experiments indicate the possibility of CSL predicting with the help of combination of the embedology and ANN approach. The technique assumes that CSL time series trace all the factors influencing on the dynamical scenario of the Caspian Sea and, hence, it does not depend on physical model of Caspian Sea dynamics. Apparently, dynamical scenarios have properties of statistical self-similarity, i.e. CSL regressions and transgressions can be observed on different time scales from decades to thousands years.

Prediction can be realized on the basis of accurate instrumental monthly time series (1830-2002). However, in this case the prediction horizon does not exceed 5 years for one prediction iteration as we have a deal with vector-based prediction.

The investigation of the instrumental and historical CSL time series have shown that these data have explicit fractal characteristics. In particular, monthly CSL data have well defined multifractal spectrum and obey to the Hurst law. The historical data regularity is entirely described by Hölder function with the mean approximately amounts to  $0.6 \pm 0.3$ . Thus, the existence of fractal properties of the CSL data has allowed application of the fractal approximation technique to process decennial historical data and to obtain annual ones. Consequently, this synthesized time series has been used for long-term CSL predictions up to 30-40 years.

## 5 Acknowledgments

We thank Dr.V. V. Golubtsov and Dr.V. I. Lee for the last Caspian Sea level data. The support from INTAS grant number 2001-0550 is gratefully acknowledged.

## References

1. S.I. Varushchenko, A. N. Varushchenko and R. K. Klige. Changes in the regime of the Caspian Sea and closed basins in paleotime. Moscow, Nauka, (1987), (in Russian)
2. Final report of Tacis Caspian Environment Programme,(2002), <http://www.dhi.dk/News/Tacis-CaspianSea/>
3. S. N. Kritskiy, D. V. Korenistov and D. Ya. Rychagov, Variation of Caspian Sea Level. Moscow, Nauka, (1975), (in Russian)
4. D. Ya. Ratkovich, *Wodnye Resursy*. **20**, 160 (1993)(in russian).
5. V. I. Naydenov, *Vestnik Russian Akademii Nauk*. **71**, 405 (2001), (in russian).
6. B. A. Apollov, *Trudy Instituta Okeanologii*. **15**, 5 (1956), (in russian).
7. V. V. Golubtsov and V. I. Lee Caspian Sea levels and the inflow of Volga River in XVI-XIX centuries, Tacis report, Caspian Environment Program, The Water Level Fluctuations Center, Almaty, March (2001), (in russian).
8. N. H. Packard, J. P. Crutchfield, R. S. Shaw, *Phys. Rev. Lett.* **45**, 712 (1980).
9. T. Sauer, J.A.Yorke, M. Casdagli, *J. Statist. Phys* **65**, 579 (1991).
10. P. E. Rapp, T. I. Schah, A. I. Mess, *Physica D* **132**, 133 (1999).
11. F. Takens, *Lecture Notes in Math.* **898**, 366 (1981).
12. E.Ott, T. Sauer and J.A.Yorke, (eds), *Coping with Chaos*,Wiley, N.Y.,1994.
13. J. P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985).
14. H. Kantz and T.Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press,Cambridge, 1997.
15. M.F. Barnsley, *Constructive approximation*. **2**, 303 (1986).
16. W.O. Cochran, J.C. Hart, P.J. Flynn, *Proc. Graphics Interface '98*, 65 (1998).
17. V. I. Naydenov and I. A. Kojevnikova, *Priroda*. **1**, 3 (2000)(in russian).
18. R. H. Riedi and I. Scheuring, *Fractals*. **5**, 153 (1997).
19. FracLab, URL: <http://fractales.inria.fr>
20. Z. R. Struzik, *Fractals*. **8**, 163 (2000).
21. V. A. Dergachev, N. G. Makarenko, L. M. Karimova and E. B. Danilkina, *Geochronometria*. **20**, 45 (2001).
22. J. D. Farmer, J. J. Sidorovich, *Phys. Rev. Lett.* **59**, 845 (1987).
23. H. D. I. Abarbanel, R. Brown and J. B. Kadtkke, *Phys. Rev. A* **41**, 1782 (1990).
24. A. N. Gorban *Appl. Math. Lett.* **11**, 45 (1998).
25. N. G. Makarenko in *Lectures on neuroinformatics*, ed. Yu. Tumenzev (MIFI, Moscow, 2003, part 1)(in russian).
26. M. Sano, Y. Sawada, *Phys. Rev. Lett.* **55**, 1082 (1985).
27. J.P. Eckmann, S.O. Kamphorst, D. Ruelle and S. Ciliberto, *Phys. Rev. A* **34**, 4971 (1986).
28. T. Tanaka, K. Aihara, and M.Taki, *Physica D* **111**, 42 (1998).
29. F. Girosi and T. Poggio, MIT AI Lab. Tech. Rep.,No 1164, Paper 45, (1989).

This page intentionally left blank

# SELF-SIMILARITY IN PLANTS: INTEGRATING MATHEMATICAL AND BIOLOGICAL PERSPECTIVES

PRZEMYSŁAW PRUSINKIEWICZ

Department of Computer Science, University of Calgary  
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4  
E-mail: pwp@cpsc.ucalgary.ca

Self-similarity is a conspicuous feature of many plants. Geometric self-similarity is commonly expressed in terms of affine transformations that map a structure into its components. Here we introduce topological self-similarity, which deals with the configurations and neighborhood relations between these components instead. The topological self-similarity of linear and branching structures is characterized in terms of recurrence systems defined within the theory of L-systems. We first review previous results, relating recurrence systems to the patterns of development that can be described using deterministic context-free L-systems. We then show that topologically self-similar structures may become geometrically self-similar if additional geometric constraints are met. This establishes a correspondence between recurrence systems and iterated function systems, which is of interest as a mathematical link between L-systems and fractals. The distinction between geometric and topological self-similarity is useful in biological applications, where topological self-similarity is more prevalent than geometric self-similarity.

## 1 Introduction

In her 1950 book, *Natural Philosophy of Plant Form*<sup>1</sup>, the eminent British botanist Agnes Arber<sup>2</sup> wrote (p. 7):

It is well to return, even at long last, to such early work as is notably rich in content, to see whether it still offers suggestions, which formerly passed unheeded because the time was not ripe for them, but which the intellectual climate would now foster. Originality is so rare in the human mind that we need to harvest it to the last gleanings.

In this paper, we follow Arber's suggestion by revisiting the notion of self-similarity in plants. Mandelbrot<sup>3</sup> defined self-similarity by referring to an underlying generative process (such as the Koch construction) as follows:

When each piece of a shape is geometrically similar to the whole, both the shape and the cascade that generates it are called self-similar.

Selected plant structures, such as the inflorescences of cauliflower and broccoli, compound fern leaves, and branching structures of trees, are often presented as canonical examples of self-similarity in the literature on fractals<sup>4</sup>. Yet aspects of self-similarity were characterized by botanists<sup>5</sup> even before the term itself was coined. One of the best such characterizations belongs to Arber herself, who wrote the following in Chapter IX ("Repetitive branching and the *Gestalt* type, with special reference to parallelism", p. 142) of her book<sup>1</sup>:

The relation to one another of a compound leaf, a simple leaf, and a mere lobe or hair, may perhaps be described as *identity-in-parallel*. A

leaflet of a compound leaf comes in, as it were, on both sides of the equation: to the compound leaf, the leaflet stands in the relation of part to whole, but it is also the equivalent to the compound leaf as the whole, though in another generation.

This quotation is interesting for several reasons. First, Arber's *identity-in-parallel* clearly anticipated the notion of self-similarity in a botanical context. Second, Arber referred figuratively to an equation, in which a form, "though in another generation", would appear on both sides. Third, Arber did not imply that this equation must necessarily have a geometric character. This leads us to the key question considered in the present paper:

How can the equation anticipated by Arber be formulated in mathematical terms?

We first examine iterated function systems as one possible interpretation of Arber's identity-in-parallel equation, and point to the botanical inadequacy of this interpretation (Section 2). We then show that Arber's identity-in-parallel can also be formalized in a different way, at the level of plant topology rather than geometry. This *topological self-similarity* can be expressed using *recurrence systems*<sup>6,7</sup> (Section 3) and their variant, *catenative formulas*<sup>8</sup> (Section 4), both of which have been defined within the theory of *L-systems*<sup>9,10</sup>. A method for conceptualizing and visualizing recurrence systems and catenative formulas makes use of data flow networks (Section 5). Recurrence systems may describe both linear and branching structures, which makes them well suited to characterize self-similarities in plant architecture (Section 6). Furthermore, with an appropriate geometric interpretation, recurrence systems may yield forms that are self-similar in both the topological and geometric sense (Section 7).

## 2 Identity-in-parallel and iterated function systems

One obvious candidate for the identity-in-parallel equation is the global characterization of fractals<sup>11,12,13</sup>, defined by the equation:

$$A = \bigcup_{i=1}^m T_i(A). \quad (1)$$

Here, the self-similar form  $A$  is the *attractor* of the set  $\{T_1, T_2, \dots, T_m\}$  of contracting transformations (usually similarities or affine transformations). This set is referred to as an *iterated function system* (IFS). Consistent with Arber's description, the attractor  $A$  appears on both sides of Equation 1. Furthermore, assuming that the initial structure  $A^{(0)}$  is given, this equation can easily be extended to a sequence of "generations":

$$A^{(n)} = \bigcup_{i=1}^m T_i(A^{(n-1)}); \quad n = 1, 2, 3, \dots \quad (2)$$

The above definitions of an IFS and its attractor can be extended to cases where different parts of a form are mapped into each other, instead of the whole form being

mapped into its own parts. These cases are captured by the closely related notions of recurrent IFS<sup>14</sup>, controlled IFS<sup>15</sup>, and language-restricted IFS<sup>16,17</sup>. In all cases, the different parts  $A_1, A_2, \dots, A_z$  of the attractor satisfy the set of equations:

$$A_j = \bigcup_{i=1}^{m_j} T_{ji}(A_{ji}); \quad j = 1, 2, \dots, z \quad (3)$$

where  $z$  is the total number of parts, and  $m_j$  is the number of transformations  $T_{ji}$  that map properly re-indexed parts  $A_1, A_2, \dots, A_z$  into part  $A_j$ . Unfortunately, even with these extensions, iterated function systems do not adequately characterize the self-similarity of plants. On one hand, IFS are insufficiently constrained: a small change in parameter values can change an attractor representing a branching structure into a set of unconnected points or segments, thus violating fundamental properties of the structures being modeled. On the other hand, IFS are too constrained: they impose a strict geometric correspondence between the form and its parts. Such correspondence is not frequently found in real plants, which is why plant-like structures generated using IFS are confined to the small set of examples that appear repetitively in the literature.

### 3 L-systems, recurrence systems, and self-similarity

In this section, we consider identity-in-parallel and self-similarity from a different perspective, focused on topology (the arrangement of components in a structure) rather than geometry. This approach is rooted in the theory of L-systems<sup>9,10,15</sup>. After background definitions, we first review a theorem linking DOL-systems (deterministic context-free L-systems) to recurrence relations between the generated strings of symbols<sup>6,7</sup>. We call these relations *recurrence systems* in a slight modification of the original definition<sup>6</sup> of this term. We then show that the recurrence systems closely correspond to Arber's notion of identity-in-parallel, and thus constitute a description of self-similarity.

**Definition 1** (from<sup>15</sup>). Let  $V$  denote a set of symbols called an alphabet,  $V^*$  the set of all words (strings of symbols) over  $V$ , and  $V^+$  the set of all nonempty words over  $V$ . A *DOL-system* is an ordered triplet  $\mathcal{G} = \langle V, \omega, P \rangle$  where  $V$  is the *alphabet* of the system,  $\omega \in V^+$  is a nonempty word called the *axiom* and  $P : V \rightarrow V^*$  is a finite *set of productions*. A production  $(a, \chi) \in P$  is written as  $a \rightarrow \chi$ . The letter  $a$  and the word  $\chi$  are called the *predecessor* and the *successor* of this production, respectively.

**Definition 2** (from<sup>15</sup>). Let  $\mu = a_1 \dots a_m$  be an arbitrary word over  $V$ . The word  $\nu = \chi_1 \dots \chi_m \in V^*$  is *directly derived* from (or *generated by*)  $\mu$ , noted  $\mu \Rightarrow \nu$ , if and only if  $a_i \rightarrow \chi_i$  for all  $i = 1, \dots, m$ . A word  $\nu$  is generated by  $\mathcal{G}$  in a derivation of length  $n$  if there exists a *developmental sequence* of words  $\mu_0, \mu_1, \dots, \mu_n$  such that  $\mu_0 = \omega$ ,  $\mu_n = \nu$  and  $\mu_0 \Rightarrow \mu_1 \Rightarrow \dots \Rightarrow \mu_n$ .

**Theorem 1** (from<sup>6</sup>). Consider a DOL-system  $\mathcal{G} = \langle V, \omega, P \rangle$ , and for each  $a \in V$

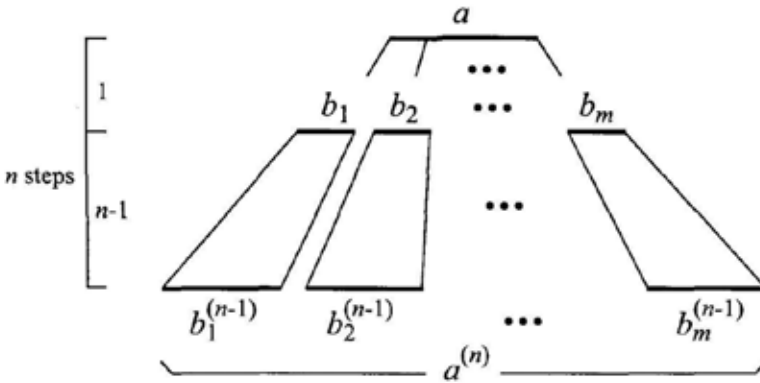


Figure 1: Illustration of the proof of Theorem 1

and  $n \geq 0$  denote by  $a^{(n)}$  the word derived from  $a$  in a derivation of length  $n$ :

$$a \xRightarrow{n} a^{(n)}. \quad (4)$$

If  $a \rightarrow b_1 b_2 \dots b_m$  is a production in  $\mathcal{G}$ , then for any  $n \geq 1$  the word  $a^{(n)}$  satisfies the recurrence formula:

$$a^{(n)} = b_1^{(n-1)} b_2^{(n-1)} \dots b_m^{(n-1)}. \quad (5)$$

*Proof.* We decompose the derivation  $a \xRightarrow{n} a^{(n)}$  into the first step and the remaining  $n - 1$  steps (Figure 1):

$$a \xRightarrow{1} b_1 b_2 \dots b_m \xRightarrow{n-1} b_1^{(n-1)} b_2^{(n-1)} \dots b_m^{(n-1)}. \quad (6)$$

Thus,  $a^{(n)} = b_1^{(n-1)} b_2^{(n-1)} \dots b_m^{(n-1)} \quad \square$ .

**Definition 3.** Given a DOL-system  $\mathcal{G} = \langle V, \omega, P \rangle$ , we call the set of the recurrence formulas given by Equation 5, along with the initial conditions  $a^{(0)} = a$  for all  $a \in V$ , the *recurrence system* associated with  $\mathcal{G}$ .









**Note.** Recurrence systems can also be specified independently of L-systems. For a formal definition and equivalence results see <sup>6,7</sup>.

**Example 1.** Consider L-system  $\mathcal{G}_1 = \langle \{a, b\}, a, P \rangle$  with productions

$$a \rightarrow ab, \quad b \rightarrow a. \quad (7)$$

This L-system can be viewed as a model of the development of a filamentous organism, with symbols  $a$  and  $b$  representing individual cells. The first production states that, over a certain time interval, a cell of type  $a$  divides into adjacent cells  $a$  and  $b$ . The second production states that, over the same time interval, a cell  $b$  changes its state into  $a$ . The above model is related to the development of the



step	L-system	recurrence system
	$\omega = a$ $a \rightarrow ab \quad b \rightarrow a$	$a^{(0)} = a \quad b^{(0)} = b$ $a^{(n)} = a^{(n-1)}b^{(n-1)} \quad b^{(n)} = a^{(n-1)}$
0		
1		
2		
3		


  
 same sequence

Figure 2: A comparison of two methods for computing sequences of words generated by L-system  $\mathcal{G}_1$ : by direct application of L-system productions (7) (left) and using recurrence relations (8) (right).

filamentous bacterium *Anabaena*, which is characterized by the unequal divisions of cells<sup>15,18</sup>. To keep the example simple, we ignore here the polarity of cells, which would determine whether a cell  $a$  divides into  $ab$  or  $ba$ .

According to Theorem 1, the recurrence system associated with  $\mathcal{G}_1$  is:

$$\begin{aligned} a^{(0)} &= a & b^{(0)} &= b \\ a^{(n)} &= a^{(n-1)}b^{(n-1)} & b^{(n)} &= a^{(n-1)} \end{aligned} \quad (8)$$

where  $n = 1, 2, 3, \dots$ . These equations provide an alternative to the usual method for generating words in an L-system, as illustrated in Figure 2. The alternative method closely corresponds to Arber's characterization of identity-in-parallel. We can make this evident by paraphrasing Arber's words quoted in the introduction to express the sample relation  $a^{(n)} = a^{(n-1)}b^{(n-1)}$ :

A substring  $a^{(n-1)}$  of the whole string  $a^{(n)}$  comes in on both sides of the equation: to the whole string, the substring stands in the relation of part to whole, but it is also the equivalent to the whole string, though in another generation.

Equating identity-in-parallel with self-similarity leads to the first, most straightforward interpretation of recurrence relations as a formal description of topological self-similarity. According to this interpretation, the words on which a recurrence system operates represent consecutive developmental stages (generations  $n$ ) of a growing structure or a set of related structures (such as the structures derived from

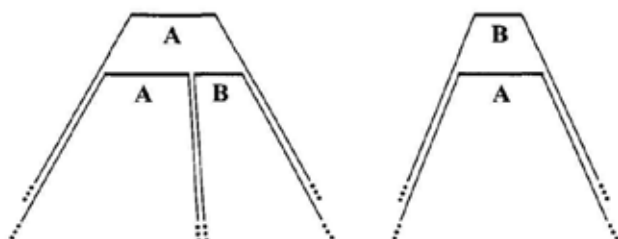


Figure 3: Self-similarity of the developmental sequences generated by L-system  $G_1$  from Example 1. Sequence **A** has itself and sequence **B** as its parts. Sequence **B** has sequence **A** as its part. This is a summary representation of the relations shown in more detail in Figure 2 (right).

$a$  and  $b$  in the *Anabaena* example). The recurrence system specifies how the younger stages (with a lower index  $n$ ) can be combined to produce the older stages. The pattern of the recurrence relations is independent of the age (generation step)  $n$ : for all  $n$  greater than some minimum value, different developmental stages of the same sequence are related to each other in the same manner.

The second interpretation of recurrence systems as a formalization of topological self-similarity deals with entire infinite developmental sequences, rather than individual words. To see this, let  $\{\mu_0, \mu_1, \mu_2, \dots\}$  and  $\{\nu_0, \nu_1, \nu_2, \dots\}$  denote two developmental sequences over some alphabet  $V$ , and  $\mathcal{N}$  and  $\circ$  denote operations on these sequences defined as follows:

$$\mathcal{N}(\xi, \{\mu_0, \mu_1, \mu_2, \dots\}) = \{\xi, \mu_0, \mu_1, \mu_2, \dots\}, \quad (9)$$

$$\{\mu_0, \mu_1, \mu_2, \dots\} \circ \{\nu_0, \nu_1, \nu_2, \dots\} = \{\mu_0\nu_0, \mu_1\nu_1, \mu_2\nu_2, \dots\}. \quad (10)$$

Thus, the  $\mathcal{N}$  operator represents a unit delay of a developmental sequence, with the initial element replaced by a given word  $\xi$ . The  $\circ$  operator represents concatenation of developmental sequences, defined as the sequence resulting from pairwise concatenation of corresponding strings in the argument sequences. By applying these definitions to the sequences

$$\mathbf{A} = \{a^{(0)}, a^{(1)}, a^{(2)}, \dots\}, \quad \mathbf{B} = \{b^{(0)}, b^{(1)}, b^{(2)}, \dots\}, \quad (11)$$

defined by the recurrence system (8), we obtain:

$$\mathbf{A} = \mathcal{N}(a, \mathbf{A} \circ \mathbf{B}), \quad \mathbf{B} = \mathcal{N}(b, \mathbf{A}). \quad (12)$$

Thus, the developmental sequences **A** and **B** include themselves as subsequences, as illustrated in Figure 3. This self-similarity of developmental sequences corresponds to the self-similarity of “cascades” in Mandelbrot’s definition quoted in the introduction.

The third interpretation of recurrence systems exposes self-similarity within individual words. To see this, let us formally interpret productions of an L-system  $G$  as local zooming operations. A derivation step  $\mu_i \Rightarrow \mu_{i+1}$  then represents the effect of globally zooming into the sequence  $\mu_i$  or, conversely, zooming out of the sequence  $\mu_{i+1}$ . Returning to Example 1 and Figure 2, a pair of words  $\{a^{(n)}, b^{(n)}\}$

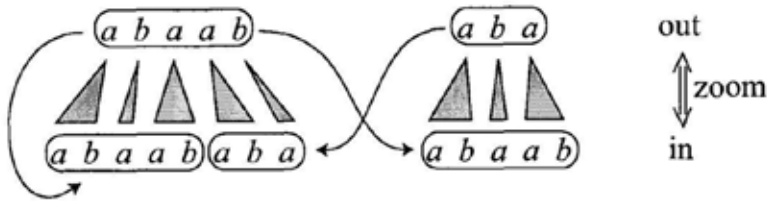


Figure 4: Self-similarity of the words generated by L-system  $\mathcal{G}_1$  from Example 1. Productions and their inverses are formally treated as scaling operations.

with  $n \geq 1$  is then equal to a combination of “scaled” versions of these words themselves, and in this sense these words are self-similar (Figure 4).

Independent of the interpretation used, Theorem 1 shows that recurrence relations are a mathematical consequence of any developmental process that can be modeled with a DOL-system. This is important from the biological perspective, because it explains why repetitive developmental processes in plants lead to topological self-similarity in developmental sequences and the resulting structures.

#### 4 Catenative formulas

An important variant of recurrence relations are *catenative formulas*<sup>7,8</sup>, which have the form

$$a^{(n)} = a^{(n-d_1)} a^{(n-d_2)} \dots a^{(n-d_m)}. \quad (13)$$

Here  $n$  is greater than some minimum value, and delays  $d_1, d_2, \dots, d_m$  are fixed positive integer numbers (i.e., they do not depend on  $n$ ). Catenative formulas express older words as combinations of younger words from the same developmental sequence<sup>a</sup>. A detailed account of the application of this concept to the description of filamentous organism (algae) is presented in<sup>19</sup>.

Referring once again to L-system  $\mathcal{G}_1$  from Example 1 and the related recurrence system (8), we observe that  $b^{(n-1)} = a^{(n-2)}$  for all  $n \geq 2$ . Using this substitution, we can express words  $a^{(n)}$  as combinations of younger (lower  $n$ ) words from the same sequence as follows:

$$a^{(-1)} = b, \quad a^{(0)} = a, \quad a^{(n)} = a^{(n-1)} a^{(n-2)}. \quad (14)$$

Thus, the developmental sequence generated by L-system  $\mathcal{G}_1$  satisfies the catenative formula (14).

Catenative formulas are important as a topological counterpart of iterated function systems; in contrast, recurrence systems are a counterpart of recurrent iterated function systems. In general, we may also consider a combination of both notions: recurrence systems that operate simultaneously on several sequences of words, and involve different delays<sup>6</sup>.

<sup>a</sup>The term commonly used in literature is *locally* catenative formula. It reflects the property that “to get a new word by catenation of some previous words it is enough to remember previous words at most  $p$  steps back in the sequence”<sup>7</sup>. However, while in this sense catenative formulas are local in time, there are not local in space. For this reason, we have dropped the reference to locality from their name.



Figure 5: A data-flow representation of the recurrence relations given by Equation 8 and illustrated in Figure 2

## 5 Data-flow network representation of recurrence systems

Iterated function systems are sometimes conceptualized as *Multiple Reduction Copy Machines* (MRCM)<sup>4</sup>, which repetitively combine reduced copies of an original figure to produce a sequence of approximations of the attractor according to Equation 2. Similarly, recurrent IFS are conceptualized using *networked MRCM*<sup>4</sup>, which operate on a set of figures according to Equation 3. In the domain of words, a related device, called a *catenation machine*, was introduced by Mavaddat<sup>20,21</sup> as a formal tool for hardware design. Below we apply catenation machines to represent recurrence systems in an intuitive, diagrammatic manner.

Referring to Figure 2, let us observe that the recurrence relations, which specify how the previously obtained words are combined into a new word, are the same at each level of a particular developmental sequence: independent of the derivation length  $n \geq 1$ . This is reflected in the repetitive pattern of lines showing “which word goes where” on the right side of Figure 2.

Instead of drawing the repetitive pattern, we can visualize recurrence relations between the old and new words using a *data-flow network*. Such a network is a directed multigraph, with the nodes labeled by symbols of the recurrence system alphabet (distinguished by a bar from the actual symbols) and arrows determined by the recurrence relations. Specifically, if  $a^{(n)} = b_1^{(n-1)} b_2^{(n-1)} \dots b_m^{(n-1)}$  is a recurrence relation then the nodes  $\bar{b}_1, \bar{b}_2, \dots, \bar{b}_m$  are connected to node  $\bar{a}$  by arrows that point to  $\bar{a}$ . These arrows are ordered in the same way as the symbols  $b_i$  in the recurrence relation. For instance, Figure 5 shows the network representing recurrence system (8) from Example 1. In general, the same letter  $b_i$  may occur in a production successor several times, and thus there may be several arrows from node  $\bar{b}_i$  to node  $\bar{a}$ , which is why the network is potentially a multigraph.

The nodes represent processing and storage elements of the network, and are capable of holding arbitrarily long words. Each node is initialized with its corresponding symbol (e.g., node  $\bar{b}_i$  initially holds symbol  $b_i$ ). The network operates in synchronous steps, in which each node concatenates the words received from its input nodes. This operation give the network its name coined by Mavaddat: the catenation machine. The network generates sequences of words according to the recurrence relations associated with a given L-system. After  $n$  steps, each node  $\bar{b}_i$  will hold the word  $b_i^{(n)}$ .

A node with a single input can be interpreted as a delay operation, which blurs the distinction between data-flow networks representing recurrence systems

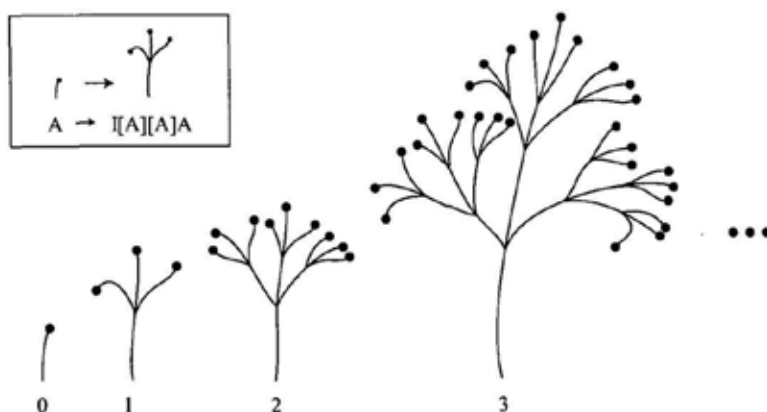


Figure 6: A sample developmental sequence with the topology described by L-system  $\mathcal{G}_2$

and catenative formulas. For example, the network shown in Figure 5 represents not only recurrence system (8), but also catenative formula (14).

Unlike recurrence relations, which can be viewed as a shorthand notation for an infinite sequence of equations corresponding to  $n = 1, 2, 3, \dots$ , the data-flow network representations are finite, with no explicit reference to index  $n$ . This reflects an essential feature of self-similarity: the repetitive character of relations between components of a structure, which in our case are represented by words. The data-flow networks capture these relations in a succinct and intuitive way, and therefore provide a convenient graphical characterization of topological self-similarities in developing structures.

## 6 Extension to branching structures

The relationship between development and recurrence systems examined in Section 3 holds not only for linear structures, but also for branching structures, which are paramount in the kingdom of plants. The extension of L-systems to branching structures makes use of the *bracketed string notation* introduced by Lindenmayer<sup>9</sup>.

**Example 2.** Let us consider a DOL-system  $\mathcal{G}_2 = \langle V, \omega, P \rangle$ , in which the alphabet  $V$  consists of four symbols: letter  $A$  denoting the apex of a branching structure, letter  $I$  denoting a branch segment, and a pair of brackets  $[, ]$  delimiting branches. The axiom is a single apex  $A$ , and the production set  $P$  has a single non-identity production:

$$A \rightarrow I[A][A]A. \quad (15)$$

(We do not explicitly specify the identity productions for the remaining symbols, such as  $I \rightarrow I$ .) According to production (15), in a given time interval the apex  $A$  creates a branching structure consisting of a segment, two lateral apices, and a terminal apex.

The above L-system generates the developmental sequence shown in Figure 6. It can be interpreted as a schematic depiction of the development of a carrot-like leaf,

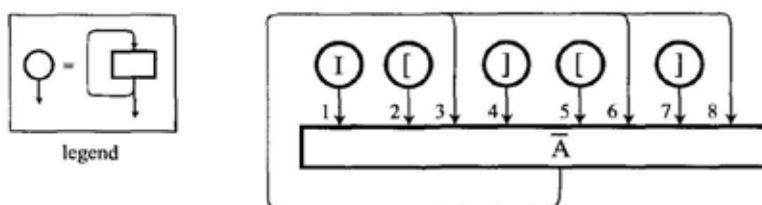


Figure 7: A data-flow representation of the recurrence relations given by Equation 16

for example. The apices are represented as circles and the segments are represented as lines. Their lengths and shapes have been chosen arbitrarily, since L-system  $\mathcal{G}_2$  only describes the branching topology of the generated structures.

In order to formally characterize the self-similar aspects of this developmental sequence, we apply Theorem 1 to construct the equivalent recurrence system:

$$A^{(0)} = A, \quad A^{(n)} = I[A^{(n-1)}][A^{(n-1)}]A^{(n-1)}. \quad (16)$$

This system is represented by the data-flow diagram in Figure 7. The structure  $A^{(n)}$ , which represents the  $n$ -th stage of the development beginning with a single apex  $A$ , can be viewed as a branching configuration of a segment  $I$  and three copies of the younger structure  $A^{(n-1)}$ . Thus, the structure  $A^{(n)}$  and the developmental sequence that generates it are topologically self-similar in all three senses of this word described in Section 3. At the same time, the structures and the developmental sequence shown in Figure 6 are not geometrically self-similar. This illustrates our thesis that topological self-similarity captures a wider class of biologically relevant phenomena, compared to geometric self-similarity.

## 7 Relation between topological and geometric self-similarity

In some cases, topologically-self-similar structures and developmental sequences can be assigned a geometric interpretation that makes them geometrically self-similar as well. This possibility is interesting from both the mathematical and the biological perspective, as it highlights the conceptual relation between topological and geometric self-similarities. The following discussion is based on a geometric extension of the L-system from Example 2.

**Example 3.** Let us consider a DOL-system  $\mathcal{G}_3 = \langle V, \omega, P \rangle$ , in which the alphabet  $V$  consists of the four symbols  $A, I, [, ]$  introduced in Example 2, and additional symbols  $+$  and  $-$  that indicate the direction of branching (to the left and to the right, respectively). The axiom is a single apex  $A$ , and the production set  $P$  has two non-identity productions:

$$A \rightarrow I[+A][-A]IA, \quad I \rightarrow II. \quad (17)$$

Compared to production (15), the first production above inserts an additional segment  $I$  between the branching point and the terminal apex  $A$ . The second production replaces each segment  $I$  with the pair  $II$  in every derivation step. These changes

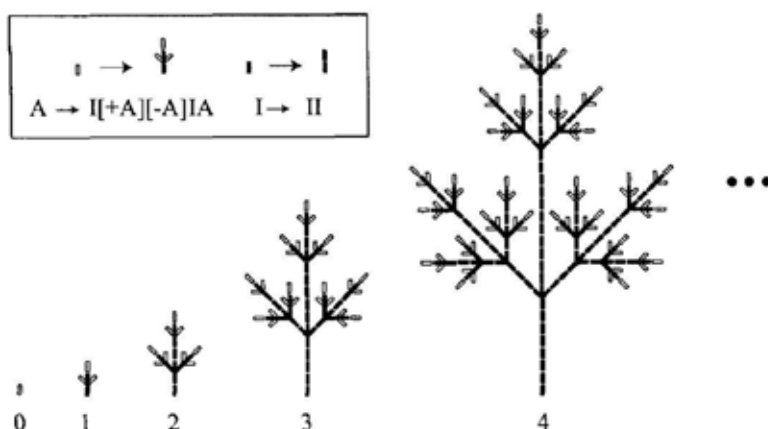


Figure 8: Developmental sequence generated by L-system  $\mathcal{G}_3$  with the turtle interpretation

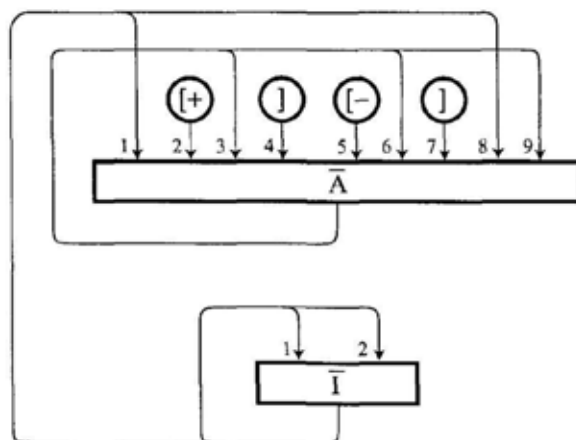


Figure 9: A data-flow representation of the recurrence relations given by Equation 18

become relevant when symbols are assigned a geometric interpretation. In our example, we use the turtle interpretation of L-system strings<sup>15,22,23</sup>. Specifically, we assume that all apices  $A$  and segments  $I$  are represented as lines of equal length, and all branching angles have the same magnitude of  $45^\circ$ . Under these assumptions, the L-system  $\mathcal{G}_3$  generates the developmental sequence shown in Figure 8.

In order to formally characterize the self-similar aspects of this developmental sequence, we apply Theorem 1 to construct the equivalent recurrence system:

$$\begin{aligned} A^{(0)} &= A & A^{(n)} &= I^{(n-1)}[+A^{(n-1)}][-A^{(n-1)}]I^{(n-1)}A^{(n-1)} \\ I^{(0)} &= I & I^{(n)} &= I^{(n-1)}I^{(n-1)} \end{aligned} \quad (18)$$

The data-flow representation of this systems is shown in Figure 9. We extract geometric information from the recurrence system (18) using the following theorem.



**Theorem 2** (from <sup>15</sup>). Consider turtle interpretation  $\mathcal{J} : V^* \rightarrow S$  as a mapping from the set  $V^*$  of words over an alphabet  $V$  into the set  $S$  of geometric figures (sets of points in the plane or in 3D space). Furthermore, let  $\mathcal{T}(\mu)$  represent the change in the turtle state (position and orientation) resulting from the interpretation of word  $\mu \in V^*$ . Then for any decomposition  $\mu_1\mu_2$  of the word  $\mu$  such that  $\mu_1, \mu_2$  do not contain unbalanced right brackets, the following holds:

$$\mathcal{J}(\mu_1\mu_2) = \mathcal{J}(\mu_1) \cup \mathcal{J}(\mu_2)\mathcal{T}(\mu_1). \quad (19)$$

According to this theorem, the turtle interpretation of the word  $\mu = \mu_1\mu_2$  is a form (set of points in two or three dimensions) equal to the union of:

1. the turtle interpretation of the word  $\mu_1$ , and
2. the turtle interpretation of the word  $\mu_2$ , repositioned by the transformation  $\mathcal{T}(\mu_1)$  that results from the interpretation of the word  $\mu_1$ .

By applying Theorem 2 to the recurrence relation given by Equation 18, we thus obtain:

$$\begin{aligned} \mathcal{J}(A^{(n)}) &= \mathcal{J}(I^{(n-1)}) \cup \\ &\quad \mathcal{J}(A^{(n-1)}) \mathcal{T}(I^{(n-1)+}) \cup \\ &\quad \mathcal{J}(A^{(n-1)}) \mathcal{T}(I^{(n-1)-}) \cup \\ &\quad \mathcal{J}(I^{(n-1)}) \mathcal{T}(I^{(n-1)}) \cup \\ &\quad \mathcal{J}(A^{(n-1)}) \mathcal{T}(I^{(n-1)}I^{(n-1)}), \\ \mathcal{J}(I^{(n)}) &= \mathcal{J}(I^{(n-1)}) \cup \mathcal{J}(I^{(n-1)}) \mathcal{T}(I^{(n-1)}). \end{aligned} \quad (20)$$

A geometric interpretation of these equations is shown in Figure 10. The generated developmental sequence exhibits an aspect of geometric self-similarity: a pair of structures  $\{A^{(n)}, I^{(n)}\}$  can be obtained by combining younger developmental stages  $\{A^{(n-1)}, I^{(n-1)}\}$  of the same structures. This combination, however, involves geometric transformations  $\mathcal{T}(I^{(n-1)})$  that change from one step  $n$  to another. In contrast, the definition of geometric self-similarity embedded in the notion of iterated function systems (Equations 1–3) postulates that transformations  $T_i$  be fixed and not depend on the iteration number  $n$ .

In the example under consideration, we can achieve this independence by changing the interpretation of symbols  $A$  and  $I$  from one derivation step to another. To this end, let us observe that all transformations  $\mathcal{T}$  in recurrence system (20) include the same term  $I^{(n-1)}$  (equal to  $I^{(3)}$  in Figure 10). According to the recurrence formulas  $I^{(0)} = I$  and  $I^{(n)} = I^{(n-1)}I^{(n-1)}$  (Equation 18), the segment  $I^{(n)}$  contains twice as many symbols  $I$  as the segment  $I^{(n-1)}$ . Thus, if the length of line segments represented by symbol  $I$  in step  $n$  is reduced by one half with respect to the lines represented by the same symbol  $I$  in step  $n-1$ , the corresponding transformations  $\mathcal{T}(I^{(n)})$  and  $\mathcal{T}(I^{(n-1)})$  will be the same, independent of  $n$ . The forms  $\mathcal{J}(A^{(n)})$  and  $\mathcal{J}(I^{(n)})$ , derived from  $A$  and  $I$  in  $n$  steps, can then be obtained by combining the forms  $\mathcal{J}(A^{(n-1)})$  and  $\mathcal{J}(I^{(n-1)})$  using transformations  $T_1, T_2, \dots, T_5$  that are independent of  $n$ . Furthermore, since the length of segments represented by symbol  $I$  (and, consistently, symbol  $A$ ) is decreased by one half between consecutive

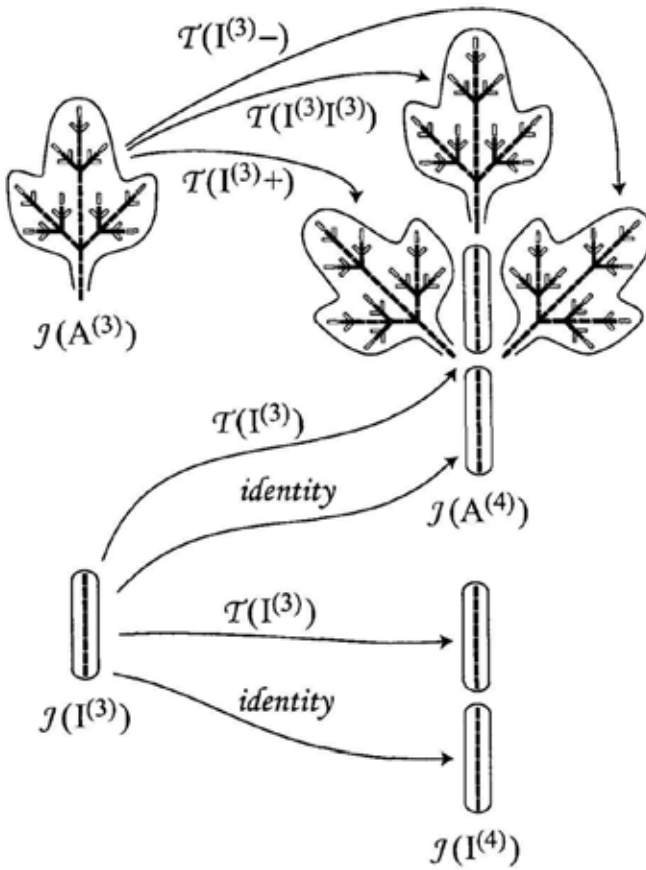


Figure 10: Geometric relations between components of the developmental sequence generated by L-system  $\mathcal{G}_3$ , as revealed by Equation 20

derivation steps, the forms  $J(A^{(n-1)})$  and  $J(I^{(n-1)})$  must also be scaled by one half before they are combined into  $J(A^{(n)})$  and  $J(I^{(n)})$  (Figure 11a).

The resulting geometric self-similarity can be characterized by a data flow diagram, in which the nodes assemble new figures by computing the set-theoretic union of figures received at the inputs, and edges represent transformations. In order to distinguish these diagrams from catenation machines, we now represent the nodes as circles (Figure 11b). The order of inputs is irrelevant, because the union of sets is a commutative operation.

The graph shown in Figure 11b represents the form of self-similarity found in recurrent iterated function systems and their variants (Section 2). A comparison of this graph with the catenation machine in Figure 9 points to the close relationship between both notions. Nevertheless, it is evident from the number of assumptions introduced in the above example that topological self-similarity yields geometric self-similarity only in special situations.

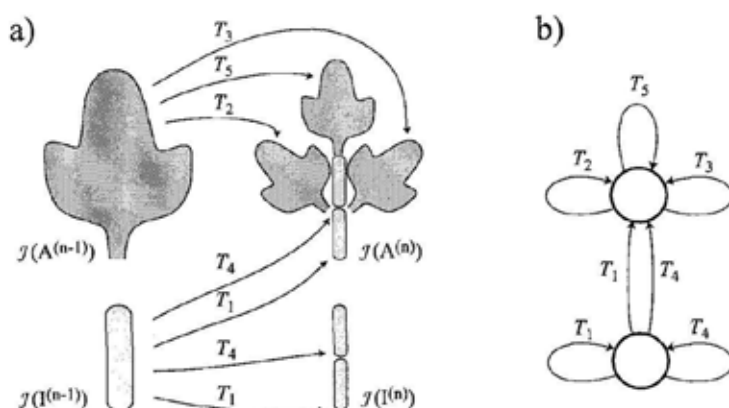


Figure 11: a) Geometric relations between components of the developmental sequence similar to that shown in Figure 10, but including an additional scaling of the components. Transformations  $T_1$  to  $T_5$  do not depend on the developmental step  $n$ . b) Control graph of the recurrent IFS associated with Figure a.

## 8 Conclusions

We revisited the notion of self-similarity in the context of plant modeling. We observed that the usual geometric self-similarity does not adequately capture self-similarity in plants, and we investigated the notion of topological self-similarity as an alternative. To this end, we revisited the notion of recurrence systems introduced in the theory of L-systems, and we concluded that they can be viewed as a formal characterization of the topological self-similarity in linear and branching structures.

Several questions are open for further research. We only considered “ordinary” (non-parametric) L-systems, and it would be interesting to extend our discussion to parametric L-systems<sup>15,24</sup> as well. It would also be interesting to investigate the general conditions under which topological self-similarity yields geometric self-similarity. More specific questions concern the relationship between delays in catenative formulas (Equation 13) and scaling transformations in the corresponding recurrent IFS. In order to firmly establish recurrence systems in the domain of fractals, it would be worthwhile to formally extend previous characterizations of the relations between L-systems, Koch constructions, and iterated function systems<sup>16,17,25,26,27</sup> to recurrence systems. The postulated structure of relationships between these formalisms is shown in Figure 12.

The last class of problems deals with applications of recurrence relations and topological self-similarity to the analysis and synthesis of real biological structures. As is well known, geometric self-similarity makes it possible to represent intricate forms using a minimum amount of data. The question is, to what extent topological self-similarity could be used in an analogous fashion, reducing the amount of data needed to describe and generate complex branching structures. In addition, it would be interesting to investigate whether the relationship between development and self-similarity could be used to infer models of plant development on the basis of self-similarities observed in mature plants.

description:	topological	geometric
local	DOL-system	Koch construction
	$\xleftrightarrow{\text{equivalence theorems}}$	$\xleftrightarrow{\text{equivalence theorems}}$
global	recurrence system	recurrent IFS
	$\xleftrightarrow{\text{geometric interpretation}}$	$\xleftrightarrow{\text{geometric interpretation}}$

Figure 12: Postulated relationships between L-systems, Koch constructions, recurrence systems, and recurrent iterated function systems

## Acknowledgments

I would like to thank Lynn Mercer and Brendan Lane for insightful comments and editorial help. This work was supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

## References

1. A. Arber. *Natural philosophy of plant form*. University Press, Cambridge, 1950.
2. R. Schmid. Agnes Arber, née Robertson (1879–1960): Fragments of her life, including her place in biology and women's studies. *Annals of Botany*, 88:1105–1128, 2001.
3. B. B. Mandelbrot. *The fractal geometry of nature*. W. H. Freeman, San Francisco, 1982.
4. H.-O. Peitgen, H. Jürgens, and D. Saupe, editors. *Chaos and fractals. New frontiers of science*. Springer-Verlag, New York, 1992.
5. W. Troll. *Die Infloreszenzen*, volume I. Gustav Fischer Verlag, Stuttgart, 1964.
6. G. T. Herman, A. Lindenmayer, and G. Rozenberg. Description of developmental languages using recurrence systems. *Mathematical Systems Theory*, 8(4):316–341, 1975.
7. G. T. Herman and G. Rozenberg. *Developmental systems and languages*. North-Holland, Amsterdam, 1975.
8. G. Rozenberg and A. Lindenmayer. Developmental systems with locally catenative formulas. *Acta Informatica*, 2:214–248, 1973.
9. A. Lindenmayer. Mathematical models for cellular interaction in development, Parts I and II. *Journal of Theoretical Biology*, 18:280–315, 1968.
10. A. Lindenmayer. Developmental systems without cellular interaction, their languages and grammars. *Journal of Theoretical Biology*, 30:455–484, 1971.
11. J. E. Hutchinson. Fractals and self-similarity. *Indiana University Journal of Mathematics*, 30(5):713–747, 1981.

12. M. F. Barnsley and S. Demko. Iterated function systems and the global construction of fractals. *Proceedings of the Royal Society of London Ser. A*, 399:243–275, 1985.
13. M. F. Barnsley. *Fractals everywhere*. Academic Press, San Diego, 1988.
14. M. F. Barnsley, J. H. Elton, and D. P. Hardin. Recurrent iterated function systems. *Constructive Approximation*, 5:3–31, 1989.
15. P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer-Verlag, New York, 1990. With J. S. Hanan, F. D. Fracchia, D. R. Fowler, M. J. M. de Boer, and L. Mercer.
16. P. Prusinkiewicz and M. Hammel. Automata, languages, and iterated function systems. In J. C. Hart and F. K. Musgrave, editors, *Fractal Modeling in 3D Computer Graphics and Imagery*, pages 115–143. ACM SIGGRAPH, 1991. Course Notes C14.
17. P. Prusinkiewicz and M. Hammel. Language-restricted iterated function systems, Koch constructions, and L-systems. In J. C. Hart, editor, *New directions for fractal modeling in computer graphics*, pages 4.1–4.14. ACM SIGGRAPH, 1994. Course Notes 13.
18. G. J. Mitchison and M. Wilcox. Rules governing cell division in *Anabaena*. *Nature*, 239:110–111, 1972.
19. H. B. Lück and J. Lück. Cell number and cell size in filamentous organisms in relation to ancestrally and positionally dependent generation times. In A. Lindenmayer and G. Rozenberg, editors, *Automata, languages, development*, pages 109–124. North-Holland, Amsterdam, 1976.
20. F. Mavaddat. Catenation machines and their relation to DTOL systems. Technical Report CS-91-26, Department of Computer Science, University of Waterloo, Waterloo, Canada, August 1991.
21. F. Mavaddat. Data-path synthesis as grammar inference. In *Proceedings of IFIP-INPG Workshop on Control Dominated Synthesis*, page 10 pp., Grenoble, France, September 1992.
22. A. L. Szilard and R. E. Quinton. An interpretation for DOL systems by computer graphics. *The Science Terrapin*, 4:8–13, 1979.
23. P. Prusinkiewicz. Graphical applications of L-systems. In *Proceedings of Graphics Interface '86 — Vision Interface '86*, pages 247–253, 1986.
24. J. S. Hanan. *Parametric L-systems and their application to the modelling and visualization of plants*. PhD thesis, University of Regina, June 1992.
25. P. Prusinkiewicz and G. Sandness. Koch curves as attractors and repellers. *IEEE Computer Graphics and Applications*, 8(6):26–40, November 1988.
26. J.C. Hart. The object instancing paradigm for linear fractal modeling. In *Graphics Interface '92*, pages 224–231. CIPS, 1992.
27. T. Ju, S. Shaefer, and R. Goldman. Recursive turtle programs and iterated affine transformations. Manuscript, Department of Computer Science, Rice University, Houston, TX, October 2003.

# COGNITIVE SCALE-FREE NETWORKS AS A MODEL FOR INTERMITTENCY IN HUMAN NATURAL LANGUAGE

PAOLO ALLEGRI

*ILC-CNR Area della Ricerca di Pisa, via Moruzzi 1, 56010 Pisa, Italy*  
*E-mail: allegrip@ilc.cnr.it*

PAOLO GRIGOLINI

*Dipartimento di Fisica, Università di Pisa and INFN, via Buonarroti 2, 56127 Pisa Italy*  
*Center for Nonlinear Science, UNT, P.O. Box 311427, Denton, Texas 76203-1427*  
*IPCF-CNR, Area della Ricerca di Pisa, via Moruzzi 1, 56010 Pisa, Italy*  
*E-mail: grigo@df.unipi.it*

LUIGI PALATELLA

*Dipartimento di Fisica, Università di Pisa and INFN, via Buonarroti 2, 56127 Pisa Italy*  
*E-mail: grigo@df.unipi.it*

We model certain features of human language complexity by means of advanced concepts borrowed from statistical mechanics. Using a time series approach, the diffusion entropy method (DE), we compute the complexity of an Italian corpus of newspapers and magazines. We find that the anomalous scaling index is compatible with a simple dynamical model, a random walk on a complex scale-free network, which is linguistically related to Saussure's *paradigms*. The model yields the famous Zipf's law in terms of the generalized central limit theorem.

## 1 Introduction

Semiotics studies linguistic signs, their meanings, and identifies the relations between signs and meanings, and among signs. The relations among signs (letters, words), are divided into two large groups, namely the syntagmatic and the paradigmatic, corresponding to what are called Saussure's dimensions<sup>1</sup>. These dimensions are analogous to physical concepts like time and space. One can grasp an understanding of them by looking at Fig. 1. The abscissa axis represents the syntagmatic dimension, while the ordinate axis represents the paradigmatic one. Along the abscissa grammatical rules pose constraints on how words follow each other. This dimension is a temporal one, with a casual order. An article (as "a" or "the"), e.g., may be followed by an adjective or a noun, but not by a verb of finite form. At a larger "time-scale", *pragmatic* constraints rule the succession of concepts, to give *logic* to the *discourse*. The other axis, on the other hand, refers to a "mental" space. The speaker has in mind a repertoire of words, divided in many categories, which can be hierarchically complex and refer to syntactical or semantic "interchangeability". Different *space*-scales of word paradigms can be associated to different levels of this hierarchy. After an article, to follow the preceding example, one can choose, at a syntactical level, among all nouns of a dictionary. However, at a deeper level, semantic constraints reduce the available words to be chosen. For instance, after "a dog" one can choose any verb, but in practice only among verbs selected by semantic constraints (a dog runs or sits, but does not read or smoke). The sentence "a dog graduates", for instance, fits paradigmatic and syntagmatic

rules behind Fig. 1, but the semantics would in general forbid the production of such a "nonsensical" sentence.

The two dimensions are therefore not quite orthogonal, and connect, e.g., at a cognitive level. The main focus of this paper is to show that this connection is in fact reproduced at all scales. We shall also show that both dimensions are *scale free* and that the complexity of linguistic structures in both dimensions can be taken into account in a unified model, which is able to explain most statistical features of human language, including, at the largest scales, the celebrated Zipf's law<sup>2</sup>.

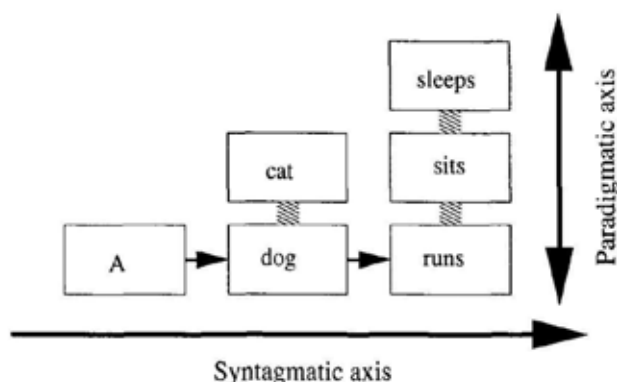


Figure 1. Saussure's dimensions. In this example the first position in the syntagmatic axis is an article, the second a noun and the third a verb in the third person.

Zipf's law relates the rank  $r$  of words to their frequency  $f$  in a corpus. Remarkably, this does not mean that the probability of a word is actually defined. In fact, a word may have a small or large frequency depending on the genre of the *corpus* (i.e. a large collection of written text) under study, and even two extremely large corpora of the same type fail in reproducing the same word frequencies. It is however remarkable that the occurrence of words is such that *for any corpus and for any natural language* a property emerge so that one finds only few frequent words and a large number of words encountered once or twice. Let us define word rank  $r$ , a property depending on the corpus adopted, as follows. One assigns rank 1 to the most frequent word, rank 2 to the second frequent one, and so on. Each word is uniquely associated to a rank, and, although this number varies from corpus to corpus, one always finds that

$$f \propto \frac{1}{r}, \quad (1)$$

This property means that *word frequencies do not tend to well-defined probabilities*. We assume that what can be defined is a "probability of having a frequency"  $P(f)$  for a randomly selected word. Operatively, one measures  $P(f)$  by counting how many words have a certain frequency  $f$ . In Section III we show that a  $P(f)$  compatible with the Zipf's law can be derived from the model proposed herein, thus



providing our model an experimental support. For our scopes, we assume a statistical mutual independence for the occurrence of different *concepts*. This hypothesis is appealing, since it means that every and each occurrence of a concept makes entropy increase, thus identifying the mathematical *information* (i.e. entropy) with the common-sense information (i.e. the occurrence of concepts). Unfortunately, a concept is not, *a priori*, a well defined quantity. Herein we assume that concepts are represented by words (or better by lemmata) or by groups of semantically similar words or lemmata<sup>a</sup>.

Because of the mutual independence among different concepts, we can extract from a single corpus as many "experiments" as the number of concepts. For each experiment we select only one concept and we mark the occurrence of the selected word or group of words corresponding to this concept. For the analysis we use the recently developed Diffusion Entropy (DE) method, which is able to identify whether a marker is a "real event", i.e. it carries maximal information, and to extract the scaling properties of the language dynamics. We show here that anomalous scaling (different from Brownian motion) is an indication of long-range correlations of the series, and that in fact these properties are well measured by the DE, even if the marker is not identified with absolute precision.

The overall dynamics, given by the flow of concepts over time, experimentally mirrors the dynamics of intermittent dynamical systems, like the Manneville's Map: These systems have long periods of quiescence followed by bursts of activity. This variability of waiting times between markers of activity is responsible for long-range correlations<sup>3</sup>.

The second aspect of the paper is the connection between space and time complexity, and its application to linguistics. We will assume that atomic concepts exist and represent nodes of a complex network, connected by arcs representing, when existing, semantic *associations* between a concept and another. We assume that our markers are actually defined as a group of neighboring nodes. We then assume that language can be produced by a random walker, "associatively" traveling from concept to concept. The scale-free properties of the network, independently measured by our research group, provides a bridge to understand the intermittent dynamics earlier described. In this unified model the network is a representation of Saussure's paradigms, whose complexity mirrors the syntagmatic one in the asymptotic limit.

## 2 DE and concepts

Let us review the DE method<sup>4,5,6,7</sup>. In synthesis, one defines a "marker" on a time sequences, and studies the probability  $p(x; t)$  of having a number  $x$  of markers in a window of length  $t$ . This statistical analysis is done by moving a window of length  $t$  along the sequences, counting how many times one finds  $x$  markers inside this window, and dividing this number by the total number  $N - t + 1$  of windows of size  $t$ , where  $N$  is the total length.

<sup>a</sup> A lemma is defined as a representative word of a class of words, having different morphological features. For instance the word "dogs" has lemma "dog", and word "sleeping" has lemma "sleep".

Having large number values for  $x$  and  $t$ , we can adopt a continuous approximation. Moreover, in the ergodic and stationary condition, a scaling relation is expected, namely

$$p(x; t) = \frac{1}{t^\delta} F\left(\frac{x - wt}{t^\delta}\right), \quad (2)$$

where  $w$  is the overall marker density,  $\delta$  is the *scaling index* and  $F$  is a function. If  $F$  is the Gauss function,  $\delta$  is the known Hurst index, and if the further condition  $\delta = 0.5$  is obeyed, then the process is said to be Poissonian, and the dynamics of  $x$  is called "Brownian motion". If this condition applies, there is no long-range memory regulating the occurrence of markers in time.

It is straightforward to show that  $S(t) = \int_{-\infty}^{\infty} dx p(x; t) \ln p(x; t)$ , namely the Shannon Information, with condition (2), leads to

$$S(t) = k + \delta \ln t, \quad (3)$$

where  $k$  is a constant. The evaluation of the slope according to which  $S$  increases with  $\ln t$  provides therefore a measure for the anomalous scaling  $\delta$ .

Let us briefly mention what we know about applying DE to time series with known long-range correlation. We construct an artificial series by letting  $\xi_i = 1$  (this means that we find the marker at the  $i$ th position), or  $\xi_i = 0$  (the  $i$ -th sign is not a marker). We then assume "informativity" for the marker (markers are then called "events"), namely that the distance between a "1" and the successive does not depend on the such previous distances. Then, if the distances  $t$  between events are distributed as

$$\psi(t) = (\mu - 1) \frac{T^{\mu-1}}{(t + T)^\mu} \quad (4)$$

( $\psi(t) \sim t^{-\mu}$  asymptotically is a sufficient condition), then the theory based on continuous-time random walk and on the generalized central-limit theorem yields for  $p(x; t)$  a truncated Lévy probability distribution function (PDF) <sup>4</sup>. DE detects the scaling  $\delta$  of the central part, namely

$$\delta = \frac{1}{\mu - 1} \text{ if } 2 < \mu < 3, \delta = 0.5 \text{ if } \mu > 3. \quad (5)$$

The condition  $2 < \mu < 3$  means long-range correlation, since for truncated Lévy PDFs asymptotically  $\langle x^2(t) \rangle - \langle x(t) \rangle^2 \propto t^{4-\mu}$  and therefore the correlation function decays as  $t^{\mu-2}$ . Note that the decay of this correlation function is non-integrable, yielding an infinite correlation time. The theory rests on a dichotomous  $\xi$ , and experimentally this means the presence or absence of a certain marker. One may, for instance look for a certain letter, so that the time is the ordinal number of the typographical characters in the text. As later shown, we have better results by looking at lemmata, where the "time" is the ordinal number of words. We shall show that, with a good choice of *semantic* markers, Eq. (4) is a good model for concepts dynamics in natural language.

Eq. (5) <sup>4</sup> rests on uncorrelated waiting times between events. This means that if two markers are separated by intervals of words of duration  $\tau_k$  (the distance in words between the  $k$ -th and the  $k+1$ -th occurrence of the marker) then  $\langle \tau_i \tau_j \rangle \propto \delta_{ij}$ , where  $\delta_{ij}$  is the Kroeneker delta. Under these conditions each event carries the same amount of information. The statistical independence between the  $\tau_k$  intervals means that the information carried by the events is maximal for a given waiting time distribution  $\psi(\tau)$ . In a linguistic jargon, we can say that if in a corpus we find a marker (e.g. a list of words) such that  $\delta \approx 1/(\mu - 1)$  then this marker is *informative* in that corpus. For didactical purposes, we shall see that certain markers, e.g. punctuation marks, are not *real events*, but are rather modeled by a Copying Mistake Map (CMM) <sup>8</sup>. This means that discourse complex dynamic is such that the punctuation marks actually carry long-range correlations, and anomalous scaling in the PDF, while the waiting times between such marks are correlated. Punctuation marks are not informative. Their complexity is just a projection of a complexity carried by “concept dynamics”.

### 2.1 The CMM and non-informative markers

The Copying Mistake Map (CMM) <sup>8</sup> is a model originally introduced to study the anomalous statistics of nucleotides dispersion in coding and non-coding DNA regions. The CMM is a combination of two sequences: We have an “original” time sequence like e.g. the long-range-correlated series earlier discussed, corresponding to the waiting time distribution (4). Then, for any  $\xi_i$  we either leave it unchanged with probability  $\epsilon$  or change it with a completely random value with probability  $1 - \epsilon$  (copying mistake).

The resulting waiting time distribution decays exponentially, since the probability of finding a 1 after a time  $t$  from the preceeding one, is given by two terms. This is because the 1 can be associated to two kinds of origin: it may be an “original” 1, or an original “zero” flipped by the copying mistake. We can write the “experimental” waiting-time distribution  $\psi_{exp}(t)$ , in terms of  $\psi_{corr}(t)$  of the mentioned long-range-correlated model (4), and of  $\psi_{rand}(t)$  of the Poissonian copying process, namely

$$\psi_{exp}(t) = \psi_{rand}(t)\Psi_{corr}(t) + \Psi_{rand}(t)\psi_{corr}(t), \quad (6)$$

where  $\Psi(t) \equiv \int_t^\infty dt' \psi(t')$  and  $\Psi_{rand}(t) \equiv \int_t^\infty dt' \psi_{rand}(t')$ , and

$$\psi_{rand}(t) = \ln \left( \frac{2}{1-\epsilon} \right) \cdot \left( \frac{2}{1-\epsilon} \right)^{-t}. \quad (7)$$

Since  $\psi_{rand}(t)$  and consequently  $\Psi_{rand}(t)$  decay as an exponential function, so it does, in the asymptotic limit,  $\psi_{exp}(t)$ . What about the DE curve? The theory predicts <sup>9</sup> a random ( $\delta = 0.5$ ) behavior for short times, a knee, and a slow transition to the totally correlated behavior. An example of CMM is given by punctuation marks in Natural Language. We choose punctuation marks as markers for an Italian corpus of newspaper and magazines, of more than 300,000 words length, called *Italian Treebank* (hereafter TB). In this experiment we look at words, and

we put a 0 for every word which is not a punctuation mark, and a 1 when we find such a mark (full stops, commas, etc.). A sentence like "Felix, the cat, sleeps!" is therefore transformed into "0 1 0 0 1 0 1".

Fig. 2 shows that this markers lead to a time series with all the earlier exposed features of a CMM. This means that the waiting times  $\tau_k$  are correlated, and therefore punctuation marks are not events. Notice however that an asymptotic anomalous  $\delta$  is detected by DE, and therefore there is a long-range correlation in the text, which may be carried by some other more informative marker.

## 2.2 Concepts as informative markers

More experiments, not reported here, show that the CMM behavior is typical for many characters, and are shared by all the letters of the alphabet, with a  $\delta \approx 0.6$ . Passing from a "phonetic" (in Italian we can assume that alphabetic characters mirror the phonetic) to a morpho-syntactic level is linguistically interesting. To do so, a text has to be lemmatized and tagged with respect to its part of speech. After this procedure we can identify as a marker the occurrence of a certain part of speech (e.g. article, adverb, adjective, verb, noun, preposition, numeral, punctuation etc.). For instance, the sentence "Felix, the cat, sleeps!" is now transformed into "N P R N P V P", where N, P, R and V stand for nouns, punctuation, article and verb. If we select the occurrence of verb as a marker, then we have "0 0 0 0 0 1 0". Fig 3 shows the result of this experiment for verbs and for numerals. We notice that we have a similar behavior for the DE, and a completely different behavior for the evaluation of the waiting-times distribution  $\psi(t)$  (where  $t$  is the number of words between markers). We notice that DE reveals a long-time correlation, while,  $\psi(t)$  shows an exponential truncation at long times. However, in the case of numerals we find a large transient with a slope  $\mu_{\text{numerals}} < 2$ , and therefore a non-stationary behavior. This is in fact due to the uneven distribution of numerals in the corpus, since they are encountered more often in the economic part of the Italian newspapers. However, this still unsatisfactory result for numerals reveals that this kind of markers is more informative than a phonetic one or than the presence of verbs. This is linguistically interesting, since numerals denote a part of speech, but also a "semantic class".

We are therefore led to suppose that informative markers are the ones associated with a semantically coherent class of words. This is however a problem, since every single concept is too rare in a balanced corpus (a long text with a variety of genres). The next level of our exploratory search for events is therefore to look at the occurrence of "salient words" in a specialistic text. Such a corpus has been made available as the Italian corpus relative to the European project POESIA<sup>10</sup>. POESIA is a European Union funded project whose aim is to protect children from offensive web contents, like, e.g. pornography in WWW URLs. Salient "pornographic" words were automatically extracted by comparing their frequency in an offensive corpus, with respect to the balanced TB-corpus. The definition adopted was

$$s(l) = \frac{f_{EC}(l) - f_{TB}(l)}{f_{EC}(l) + f_{TB}(l)}, \quad (8)$$

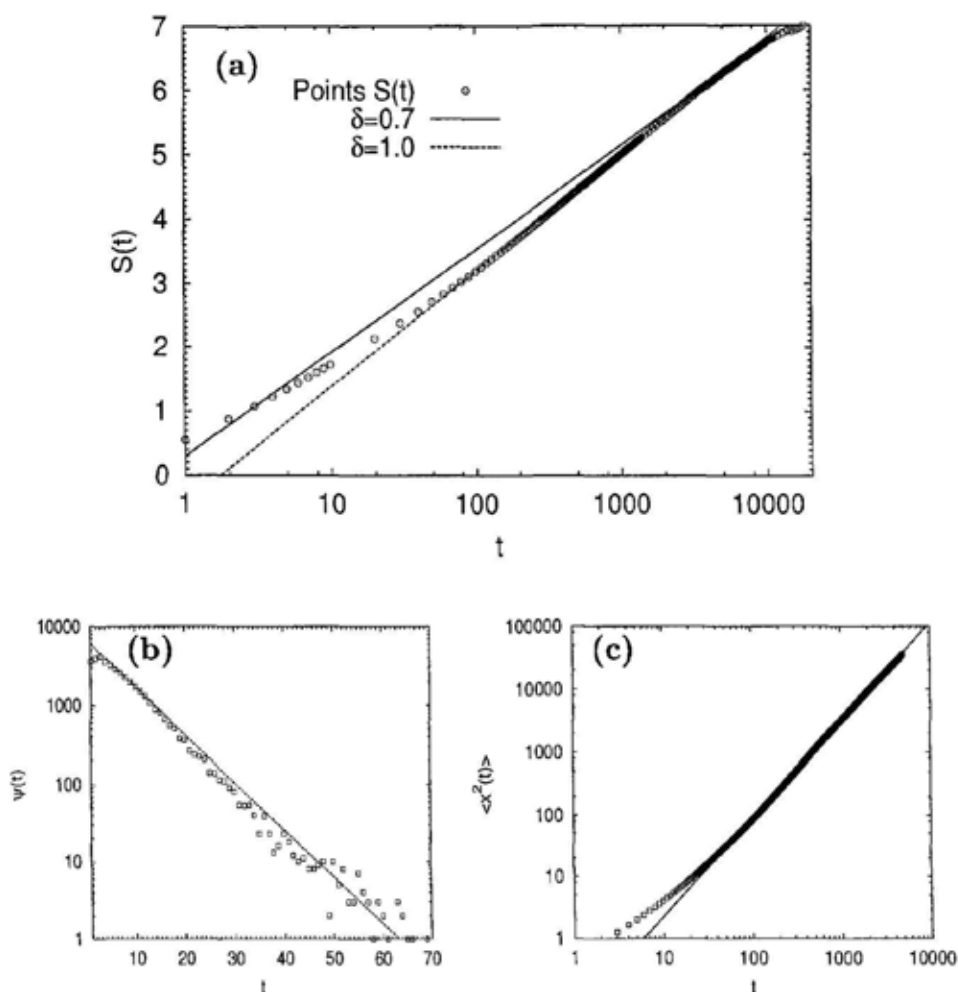


Figure 2. (a) Diffusion entropy for punctuation marks. The fit for the asymptotic limit (solid line) yields a  $\delta = 0.7$ . The dashed line marks a transient regime with  $\delta = 1$ . (b) Non-normalized distribution of waiting times for punctuation marks, namely counts of waiting times of length  $t$  between marks in TB. The expression for the dashed line fit is  $7000 \cdot \exp(-t/7.15)$ . (c) Second moment analysis for punctuation marks. The expression for the solid line fit is  $0.06 \cdot t^{1.57}$ . Notice that  $1.57 \approx 3 - 1/0.7$ , namely the expression  $H = 3 - 1/\delta$  of Ref. 9 for Lévy processes stemming from CMM's is verified.

where  $f_{EC}(l)$  is the frequency, in the erotic corpus, of the lemma  $l$ , and  $f_{TB}(l)$  is the same property in the reference Italian corpus (Italian Treebank). Salient lemmata were automatically chosen as the 5% with the highest value of  $s$ . Notice that in this experiment all “dirty” words are not taken into consideration, because they do not appear in the reference corpus, and therefore  $s$  cannot be properly

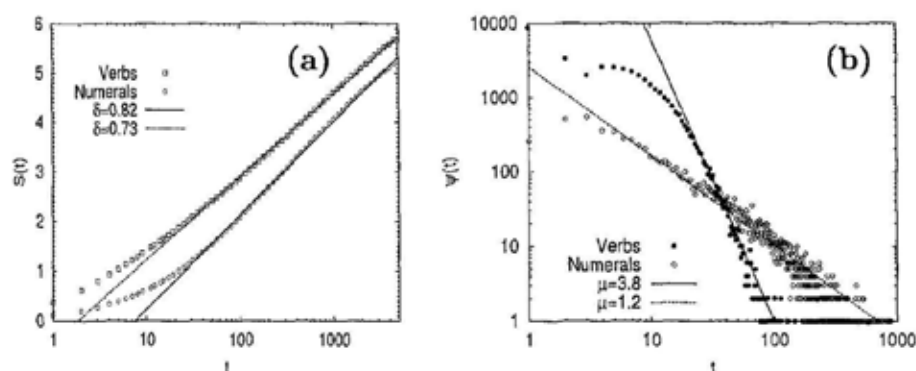


Figure 3. a) DE for verbs (squares) and numerals (circles). The dashed line is a fit for the verbs, with  $\delta = 0.73$ , while the solid line is a fit for the numerals, with  $\delta = 0.82$ . b)  $\psi(t)$  for verbs (black circles) and numerals (white circles). The dashed line is a fit for the numerals, with  $\mu = 1.2$ , while the solid line is a fit for the verbs, with  $\mu = 3.8$ .

defined. However an offensive metaphoric use of terms is in fact detected, leading to a completely new way to automatic text categorization and filter <sup>10</sup>, using a method, based on DE analysis, called CASSANDRA <sup>11</sup>.

Salient words were therefore used as markers for our analysis, as earlier described. The results are shown in Fig. 4, clearly showing that *in a specialized corpus, salient words of this genre, pass the test of informativeness*. Salient words, and plausibly words in general, are therefore distributed like markers generated by an intermittent dynamical model, with  $\mu \approx 2.1$  and, in agreement with (5),  $\delta \approx 1/(\mu - 1) = 0.91$ . We see in the next section how this behavior is plausibly connected with a topological complexity at the paradigmatic level, and in Section III we derive the Zipf's law from the resulting model.

### 3 Scale-free networks, intermittency and the Zipf's law

In this section we build a cognitive model for connecting structure and dynamics. Allegrini et al. <sup>12</sup> identified semantic classes in the Italian corpus, by looking at paradigmatic properties of interchangeability of classes of verbs with respect to classes of nouns. They defined "superclasses" of verbs and nouns as "substitutability islands", namely groups of nouns and verbs sharing the properties that in the corpus you find each verb of the class co-occurring, in a context, with each noun of the class <sup>12</sup>. This is precisely a direct application of the notion of "paradigm". Let us call  $p_v(c)$  and  $p_n(c)$ , respectively, the number of verbs or nouns belonging to a number  $c$  of classes. They found that

$$\begin{aligned} p_v(c) &\propto \frac{1}{c^{1+\eta}} \\ p_n(c) &\propto \frac{1}{c^{1+\eta}}, \end{aligned} \quad (9)$$

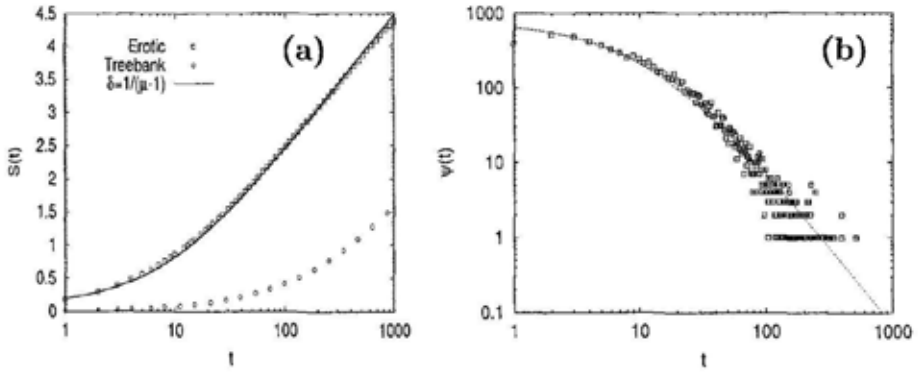


Figure 4. a) DE for salient “erotic” words for a corpus of erotic stories and offensive web pages (squares), and for the Italian reference corpus (circles). The solid line is a fit with expression  $S(t) = k + \delta \ln(t + t_0)$ , where the additional parameter  $t_0$  is added to the original Eq. (3) to take transients into account and to improve the quality of the fit, yielding  $\delta = 0.91$  b) Non-normalized waiting time distribution for salient “erotic” words for a corpus of erotic stories. The expression for dashed line fit is  $14000 \cdot (12.0 + t)^{-2.1}$ , yielding  $\mu = 2.1$ .

where  $\eta$  is a number whose absolute value is (much) smaller than 1.

On the same line, other authors<sup>13</sup> found a “small world” topology<sup>14</sup>, by looking at the number of synonyms in an English thesaurus, for each English lemma. We can therefore assume that this kind of structure is general for any language. Let us therefore imagine that the paradigmatic structure of concepts is a scale-free network and consider a random walk in this “cognitive space”. Let us make the following assumptions:

- 1) The statistical weight of the  $i$ -th node is  $\omega_i \sim c_i$ ;
  - 2) Ergodicity, and therefore that the characteristic recurrence time is  $\tau_i \sim c_i^{-1}$ ;
  - 3) The same form for all nodes,  $\psi_i(t) = (1/\tau_i)F(-t/\tau_i)$  (e.g.  $F(x) = \exp(-x)$ ).
- Now we imagine that selecting a *concept* means selecting a few neighboring nodes. This collection of nodes, due to the scale-free hypothesis, shares the same scaling properties of the complete scale free network, namely  $p(c) \sim c^{-\nu}$ . Therefore we have that

$$\psi_{concept}(t) = \sum_i \omega_i \psi_i(t) \propto \sum_i c_i^2 e^{-c_i t} \approx \int dc c^2 e^{-ct} \frac{1}{c^\nu} \sim \frac{1}{t^{3-\nu}}. \quad (10)$$

We recovered the intermittent model (4).

Let us now make the exercise of deriving the Zipf’s law  $f \propto r^{-a}$ , with  $a$  close to unity. Let us define a probability of frequency  $P(f)$

$$P(f)df = prob(r)dr \implies P(f) \sim f^{-\frac{a+1}{a}} \quad (11)$$

Next, let us notice that  $P(f)$  must be a stable distribution. In fact, the Zipf’s law is valid for every corpus. In particular if it is valid for corpus  $A$  and for corpus  $B$ ,



it is valid also for the corpus  $A + B$  where  $+$  means the concatenation of corpora. If we continue with concatenating we will have a corpus

$$\text{Total Corpus} = \text{Corpus A} + \text{Corpus B} + \dots$$

and we write the frequency of a word in the total corpus,  $f_{tot}$  is written in terms of the single frequencies  $f_1, f_2, \dots$ , and total lengths  $N_1, N_2, \dots$  of the single corpora

$$f_{tot} = \frac{f_1 + f_2 + \dots}{N_1 + N_2 + \dots} = \frac{1}{\sum_i N_i} \sum_i f_i, \quad (12)$$

i.e., the Generalized Central Limit Theorem<sup>15</sup> applies. This means that the probability of frequency  $P(f)$  is a Lévy  $\alpha$ -stable distribution. This probability of finding  $f$  occurrences of a word in a corpus of a given length can be identified with  $p(x; t)$  of Section II, if we take into consideration the parameter  $t$ . We have earlier noticed that  $p(x; t)$  in language is Lévy process, with  $\delta \sim 1$ , and therefore with a tail  $P(f) \sim f^{-2}$ . In other words through (11) we recover (1) i.e. the Zipf's law.

#### 4 Conclusions

In this paper we have shown that a cognitive process governing human language may be identified, and that it has a complexity both in the syntagmatic and in the paradigmatic axis. The scaling properties of both axes are related to each other, and are reflected by the celebrated Zipf's law. This study was conducted using Italian written corpora, but decades of studies on the generality of the Zipf's law lead us to suppose that our results are language independent, and that the language complexity that we are revealing is genuine and important. In fact, for any concept, we have a scaling index associated with an intermittent dynamical model that rests at the border between ergodicity and non-ergodicity, since the Zipf's law is theoretically consistent with  $\delta = 1$ . Moreover, in a specialistic test we see a tendency to drift, for salient words, towards ergodicity ( $\delta \approx 0.91$  in the reported experiment). This behavior can be interpreted as the balance between two opposite needs for human language, namely *learnability*, i.e. the possibility for a child to learn a language by examples, and *variability*, to explore an infinite cognitive space.

We propose as a future work to study language complexity in children during learning years, and in psychopathological subjects. We imagine, if the theory presented herein is validated by more extensive work, that the simple study of the individual Zipf's laws can provide a reasonable non-invasive diagnostic method for certain mental diseases.

From a Language Engineering point of view, this study provides a theoretical background for a completely new strategy of automatic text categorization. A prototype is being implemented as a semantic filter<sup>10</sup>. We think that the proposed test for informativeness for a set of markers can also be important for many exploratory studies in time series analysis. For instance, it may become important to identify crucial semantic markers in a flow of data.

## References

1. K. Silverman, *The Subject of Semiotics*, (Oxford Univ. Press, New York, 1983).
2. GK Zipf, *Psycho-Biology of Languages* (MIT Press, Cambridge MA, 1965)
3. P. Manneville, *J. Physique* **41**, 1235 (1980).
4. P. Grigolini, L. Palatella, G. Raffaelli, *Fractals* **9**, 439 (2001).
5. P. Allegrini, R. Balocchi, S. Chillemi, P. Grigolini, P. Hamilton, R. Maestri, L. Palatella, G. Raffaelli, *Phys. Rev. E* **67**, 062901 (2003).
6. P. Allegrini, V. Benci, P. Grigolini, P. Hamilton, M. Ignaccolo, G. Menconi, L. Palatella, G. Raffaelli, N. Scafetta, M. Virgilio, J. Yang, *Chaos, Solitons & Fractals* **15**, 517 (2003).
7. S. M. Mega, P. Allegrini, P. Grigolini V. Latora, L. Palatella, A. Rapisarda, S. Vinciguerra, *Phys. Rev. Lett.* **90**, 188501 (2003).
8. P. Allegrini, M. Barbi, P. Grigolini and B. J. West, *Phys. Rev. E* **52**, 5281 (1995); P. Allegrini, M. Buiatti, P. Grigolini and B. J. West, *Phys. Rev. E* **58**, 3640 (1998).
9. N. Scafetta, V. Latora, P. Grigolini, *Phys. Lett. A* **299**, 565 (2002); N. Scafetta, V. Latora, P. Grigolini, *Phys. Rev. E* **66**, 031906 (2002).
10. Visit URL <http://www.poesia-filter.org>, for all information about Poesia (Public Open-source Environment for a Safer Internet Access), European Project Number IAP 2117/27572 (2002), and the open-source Poesia filter.
11. P. Allegrini, P. Grigolini, L. Palatella, G. Raffaelli, M. Virgilio, in *Emergent Nature*, ed M.M. Novak (World Scientific, Singapore, 2002).
12. P. Allegrini, S. Montemagni and V. Pirrelli, in *COLING Proceedings*, (ed. COLING, Saarbruecken 2000); P. Allegrini, S. Montemagni, V. Pirrelli, *Rivista di Linguistica Computazionale*, in press.
13. A.E. Motter, A.P.S. de Moura, Y.-C. Lai, and P. Dasgupta, *Phys. Rev. E* **9**, 065102 (2002).
14. D.J. Watts and S.H. Strogatz, *Nature* **393**, 440 (1998); A.-L. Barabási, *Linked, The New Science of Networks*, (Perseus Publishing, Cambridge, MA, 2002).
15. W. Feller, *An introduction to probability theory and its applications*, vol. 1, (Wiley, New York, 1971).

This page intentionally left blank

# THE COMPLEXITY OF BIOLOGICAL AGEING

DIETRICH STAUFFER

*Institute for Theoretical Physics, Cologne University, D-50923 Köln, Euroland*

The present review deals with the computer simulation of biological ageing as well as its demographic consequences for industrialized societies.

## 1 Introduction

Life usually ends with death, and ageing is defined here by the increase of the mortality rate with increasing age. Merryll Streep and Goldie Hawn showed in the movie "Death becomes her" the consequences of an elixir giving us eternal life. The present review instead deals with the consequences present foreseeable trends have on the demography of developed countries, and with the biological reasons of ageing.

## 2 Demography

The mortality rate  $q$  is the fraction of people of age  $x$  who die within the next time unit, i.e. before they reach age  $x + 1$ :  $q(x) = [S(x+1) - S(x)]/S(x)$ . Here,  $S(x)$  is the probability to survive from birth to age  $x$ . This quantity  $q$  can by definition not increase beyond  $q = 1$ . It depends on the time unit which is typically a year for humans and a day for flies and worms. A better quantity, which can increase beyond unity, is the derivative  $\mu$  for infinitely small time steps instead of discrete time steps,

$$\mu(x) = -d \ln S(x)/dx, \quad (1)$$

called here the mortality function (also denoted as hazard rate or force of mortality<sup>1</sup>). If life tables with yearly units are published, then  $\mu$  can be approximated through

$$\mu(x + 1/2) \simeq \ln S(x) - \ln S(x + 1) \quad (2)$$

which also can increase beyond unity. The astronomer Halley tried about three centuries ago to find some laws governing human mortality, but only in the 19th century Gompertz found the law which is valid when childhood diseases are overcome, Fig. 1:

$$\mu \propto \exp(bx) \quad (3)$$

with an empirical parameter  $b$  increasing for humans over the centuries. It holds also approximately for many animals in protected environments like zoos and laboratories<sup>2</sup>, while in the wild many animals are eaten by predators before they reach their genetically possible age. Below the age of 25 years for humans, deviations are seen: The mortality (rate or function) is high at birth (most human embryos die before birth), then sinks to a broad minimum between 5 and 12 years, and only then increases monotonically up to old age.

From recent life tables of Swedish men, with Gompertz law for 30 to 90 years, and childhood diseases

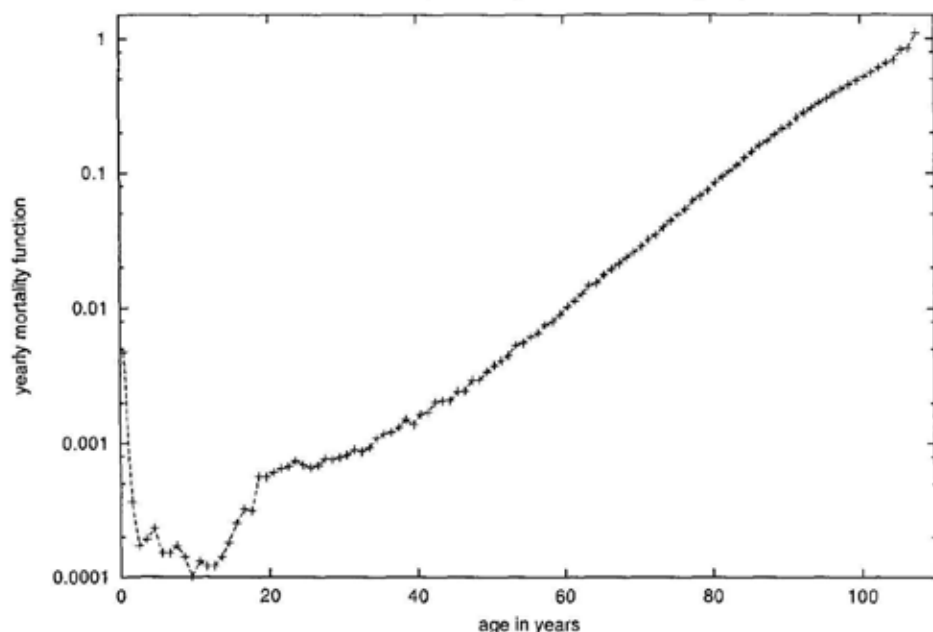


Figure 1. Gompertz law ad middle age, childhood deviations, and possible downward deviation from Gompertz law for centenarians among Swedish men. From life tables 1993-1997 of the central statistical bureau in Sweden: Ewa.Eriksson@scb.se (1999).

For humans, the last two centuries have seen a doubling of the life expectancy at birth,  $\int S(x)dx = \int x\mu(x)S(x)dx$ , from 40 to 80 years. Figure 2 shows, for Swedish women since 1750, both the prolongation of life and the increasing reliability. (We did not all die at age 40 in earlier centuries. If half of the babies die in their first year, and the other half lives until 80, then the life expectancy is 40 years.) We see clearly that the life expectancy no longer increases as fast as in the first half of the 20th century, but the present lower rate of increase showed no sign of further reduction during the last decades. Figure 3 shows that this increase of life expectancy comes not only from the reduction of child mortality but also from an increase of the remaining life expectancy at age 65.

Much more recent is the reduction of the birth rate (number of children born per women during her lifetime) below the replacement level of about 2.1. The two German states started this around 1970, due to "the pill", and in West Germany the birth rate scattered about 1.4 in the last three decades. In France it is higher, in Spain and Italy lower. World War II was started by Nazi Germany with the excuse of "Volk ohne Raum", than the Germans needed more living space; so a reduction of the native population (enlarged by immigration) did not seem bad around 1970. In the meantime, however, the reduction in the number of young people coupled with the increase in the number of old people is seen as a threat to the usual way in which you should finance my retirement. If the strongest age cohort in the year



Figure 2. Life expectancy for Swedish women for 250 calendar years. From Wilmoth's Berkeley Mortality Data Base.

2030 will be people at age 70, we can hardly afford an average retirement (healthy Germans) at 62.

Thus we <sup>6</sup> (and others) predicted the future ratio of pensioners to working-age people, assuming that after the year 2011 the retirement age is increased from 62 years by 0.6 years for every year by which the life expectancy at birth increases, and that starting in 2005 an immigration of 0.38 percent per year (of the population) of people aged 6 to 40 years keeps the total population stable. We see that the dangerous peak around the year 2030 is followed by a plateau in this ratio. (Working was assumed to start at age 20.)

In a comparison of life tables for different countries and different countries, a certain degree of universality was found for the human Gompertz law  $\mu \propto \exp(bx)$ : The mortality for centenarians was about the same <sup>3,4,5</sup>. Thus

$$\mu \simeq 7b \cdot \exp[b(x - X)] \quad (4)$$

with a characteristic age of  $X \simeq 103$  years for the whole human species, while  $b \sim 0.1$  increases with time. (For  $x < X$  the differences between  $q$  and  $\mu$  are quite small.) Moreover, Azbel <sup>7</sup> found (with some deviations) a universality even for the younger ages where Gompertz is invalid, Figure 1. He found the mortality  $q_x(c, t)$  at age  $x$  (for country  $c$  and calendar year  $t$ ) to be a universal function  $f_x$  of infant mortality  $q_0 = q_{x=0}(c, t)$  and age  $x$ :  $q_x(c, t) = f_x(q_0)$ . The function  $f_x$  no longer depends explicitly on  $c$  and  $t$ , in contrast to  $q_x$ . Thus if country A has at present

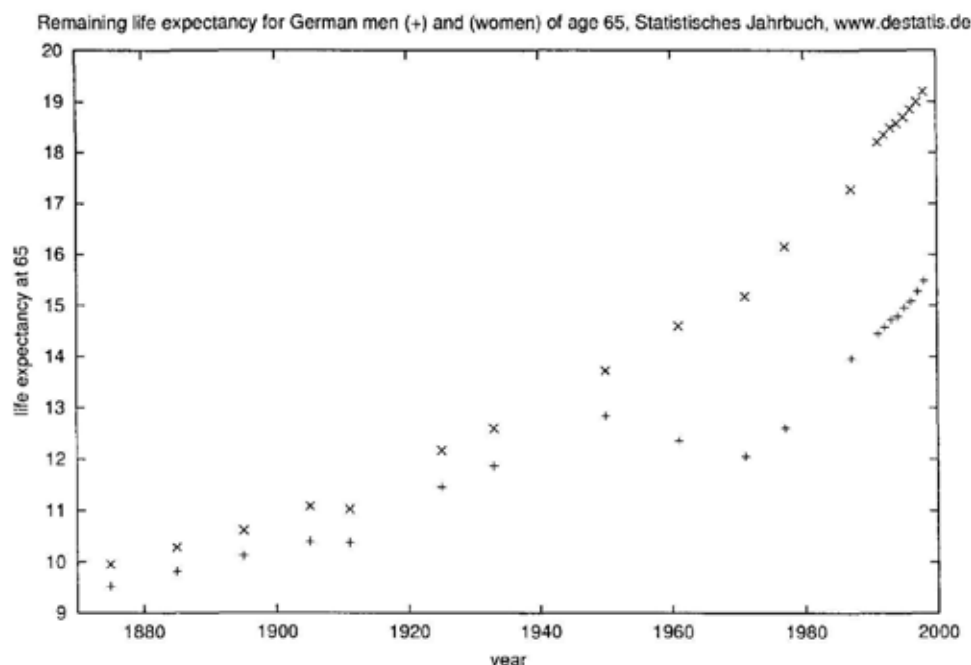


Figure 3. Remaining life expectancy at the German legal retirement age of 65.

a known mortality function of age, then another country B roughly has the same mortality function if we change the calendar year  $t$  such that the infant mortality in B at time  $t$  agrees with the present infant mortality in A. If the Gompertz law would be valid for all ages, Azbel's universality<sup>7</sup> would already follow from Eq.(4) since it contains only one free parameter  $b$  for all human societies. These universality laws suggest that extrapolations like Fig. 4 may, with some shift in time  $t$ , be valid also for developing countries, if they do not take early action to keep the birth rate near the replacement level of 2.1 or whatever else is needed to offset deaths and net emigration.

A decade ago, mortality maxima were observed<sup>8,9</sup> for flies. Have they found the fountain of youth such that we get healthier again with increasing age? Humans at least, Fig. 1, do not show such maxima in reliable statistics, though USA data published in the 1990 showed them. (Reliability seems to increase from USA to Western Europe and from there to Sweden.) Perhaps above 110 years human mortality reaches a plateau<sup>1,10</sup>. But a comparison of Figures 3 and 4 in<sup>10</sup> shows that for the more reliable half of the European data, the highest claimed ages were appreciably below those of all the data: The more reliable the data are the smaller are the deviations from the Gompertz law. Perhaps for the oldest old the mortality still increases with age, but only linearly<sup>13</sup> and not exponentially: Neither acceleration nor deceleration of mortality. More arguments against mortality deceleration for humans are given elsewhere<sup>11,12</sup>.

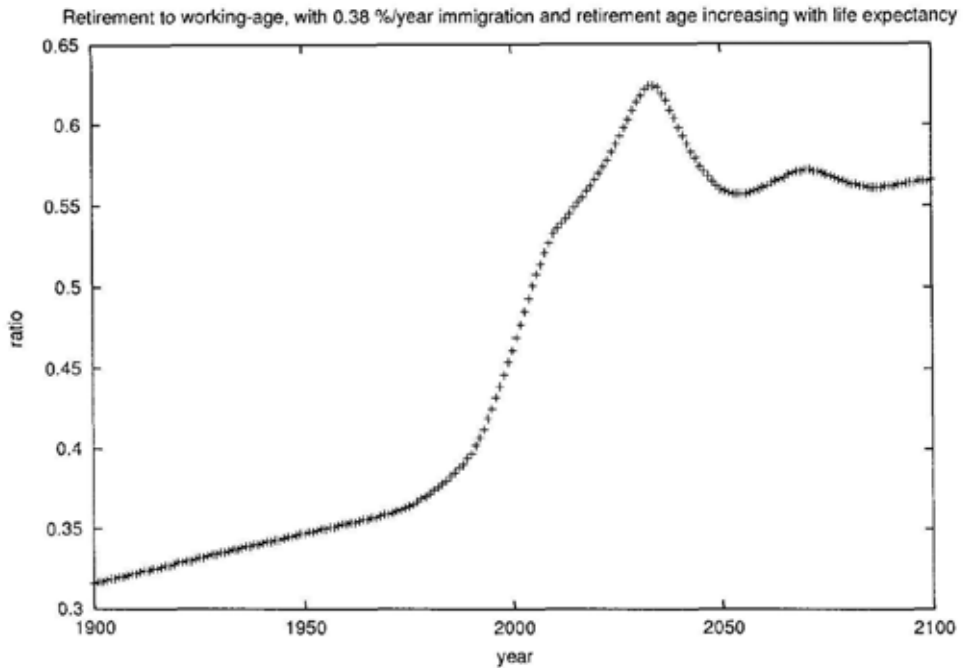


Figure 4. Ratio of people above retirement age, to working-age people older than 20 years, as predicted in Ref. 6 using extrapolated Gompertz laws.

While at present we thus should be cautious about buying fountains of youth for humans, future decades might produce genetically modified humans with longer life expectancies. The little worm *Caenorhabditis elegans* survives bad times (no food, ...) by reducing all life functions during a "dauer" state, in agreement with computer simulations<sup>14</sup>. Even flies and some mammals live longer if put on a starvation diet<sup>15</sup>. But do we want to live longer if the gained life span is spent in a coma, or in hunger? More attractive is a very old elixir of youth, red wine. According to<sup>16</sup>, the polyphenol resveratrol in red wine activates the Sir2 gene and lets yeast cells live 70 % longer. Indeed, Jeanne Calment is widely (though not universally) believed to be the oldest human being and died in 1997 at age 122 in Southern France, having drunk red wine moderately. Let me see if drinking it beyond moderation lets me beat her world record.

### 3 Why do we age?

It may be an exaggeration that there as many theories of ageing as there are ageing theorists, but nevertheless we have lots of theories. They even might all be correct, since ageing may have many causes. Also, some theories do not exclude each other, describing only different aspects of the same phenomenon.

120 years ago, Weismann suggested that we die to make place for our children



<sup>17</sup>. This is very altruistic but also very unrealistic, since of two different races of the same species the one which produces more children will win the Darwinian struggle of survival of the fittest. Those who live longer at otherwise unchanged parameters produce more children and thus win in the short term even if on long time scales they drain the environmental resources stronger and might finally destroy the ecosystem, including themselves. (If, however, longer lifetime is coupled to lower birth rates, the Weismann idea becomes viable as explained later.)

Medawar <sup>18</sup> suggested more than half a century ago the still relevant mutation accumulation theory: Bad mutations killing us in young age before we get children will die out since they are not given on to the offspring; bad mutations killing us after we produced many children are given on to future generations. Thus after some time the population should contain few hereditary diseases affecting us in young age, but many affecting us in old age. Thus the probability to die from them increases with increasing age.

A now widespread idea are oxygen radicals created by metabolism and destroying the DNA, carrier of heredity, during our life. This theory is not necessarily in contradiction with the mutation accumulation; instead it is a biochemical explanation for these mutations.

Telomeres are sections at the end of the DNA which are lost at every duplication of the cell. If the number of telomeres in this way has become too small, the cell stops dividing: Hayflick limit. A recent ageing theory <sup>19</sup> is based on these telomeres, and perhaps at the time of the conference I can present more simulations.

Longevity genes would prolong life, have been found to work for many animals, and are perhaps connected with the red wine effect mentioned at the end of section 2. Again, their existence does not contradict the other theories: A longevity gene may produce more telomerase, an enzyme which restores lost telomeres. Or it may enhance scavenging of dangerous oxygen radicals and thus reduce the mutation rate relevant for mutation accumulation.

Reliability theory <sup>4</sup> may work also for the ageing of automobiles and was connected to the Penna model (see below) in <sup>20</sup>, see also <sup>21</sup>. It assumes the organism or car to consist of  $m$  irreplaceable blocks; failure of one of these blocks causes the whole system to fail. Each block consists of many equivalent elements; the initial numbers of properly working elements within a block follow a Poisson distribution. A block fails if all its elements fail; each element ages with a constant failure rate  $1/X$ . Then Gavrilov and Gavrilova <sup>4</sup> recovered analytically the Gompertz law, Eqs.(3,4), for age  $x \ll X$  and a mortality plateau  $\mu = m/X$  for very high age  $x \gg X$ . The characteristic age  $X$  in Eq.(4), valid for all humans, then is the average lifetime of the single elements.

The following sections will report computer simulations of the mutation accumulation idea.

#### 4 Simulations of mutation accumulation

This section restricts itself to those individual-based ageing models which were investigated in papers of different groups. Historically the first are those of Partridge-Barton type <sup>22</sup>, followed by the most widely used Penna model <sup>23</sup>, while the Stauffer

model <sup>12</sup> is more of conceptual than of practical value but therefore forms our starting point.

In contrast to Weismann, we do not die to make place for our children. But if we fix the number of children, then the idea <sup>12</sup> works: The birth rate (per iteration) is assumed to be inversely proportional to the time between the minimum age of reproduction,  $x_m$ , and the genetic death age,  $x_d$ . Hereditary mutations accumulate over the generations, and each may independently change both characteristic ages  $x_m$  and  $x_d$  by one time unit. Individuals may die before Thai genetic death age from hunger etc, which is taken into account by a Verhulst death probability proportional to the population size, as in a logistic equation. Then automatically a reasonable distribution of death ages emerges and death is explained as coming from random mutations plus a trade-off between longevity and high birth rate. The catastrophic senescence of Pacific salmon, the death of Northern cod though over-fishing, the minimum population size for social animals, and the emergence of female menopause were simulated successfully <sup>24,25,26</sup>. However, in general <sup>27</sup> the mortality increases linearly with age, instead of the desired exponential Eq.(3). Also, the minimum age of reproduction is distributed among unrealistic short ages, even in a much more complicated model of a whole ecosystem <sup>28</sup>.

This trade-off between longevity and high birth rate is mentioned a lot in the biological literature. The most direct but not the only way to realize it genetically are mutations with antagonistic pleiotropy <sup>22,29</sup>: these genes have positive effects in youth and negative ones at old age.

Computer simulations of ageing started by putting fluctuations into the phenomenological model of Partridge and Barton <sup>22,30,31,32,14,33,34</sup>. Originally it assumed only three ages zero, one and two, with a juvenile survival rate  $J$  from zero to one, and an adult survival rate  $A$  from one to two. It was first thought to give unrealistic mortality functions and difficulties if generalized to many age levels, but <sup>33</sup> repaired this by slight modifications, and <sup>34</sup> included sexual reproduction in it. But the lack of an explicit genome makes it less attractive than the Penna model described now.

The Penna model is by far the most widespread method to simulate biological ageing. Most of the literature up to 1998 is cited in <sup>35</sup>, and later work up to 2000 in <sup>12</sup> for asexual and <sup>36</sup> for sexual reproduction. The genome is represented by a bit-string ( $10^1 \dots 10^3$  bits were simulated) giving bad mutations. A bit set to zero is healthy, a bit set to one means that a hereditary disease starts to reduce the health at the age to which the bit position corresponds. The first bits describe diseases starting in youth, which are rare, and the latest bits correspond to the much more frequent diseases at old age. Three active diseases kill, and so does a Verhulst death probability proportional to the total population size. Each time interval, every individual above the minimum age of reproduction produces offspring which differs from the parent by a random mutation of the bit-string genome. In the sexual version with recombination of the two bit-strings of the genome, dominant mutations affect the health already if only one bit-string is mutated, while recessive mutations become dangerous only if both bit-strings are mutated. Bigamy with three bit-strings was discussed in <sup>37</sup>.

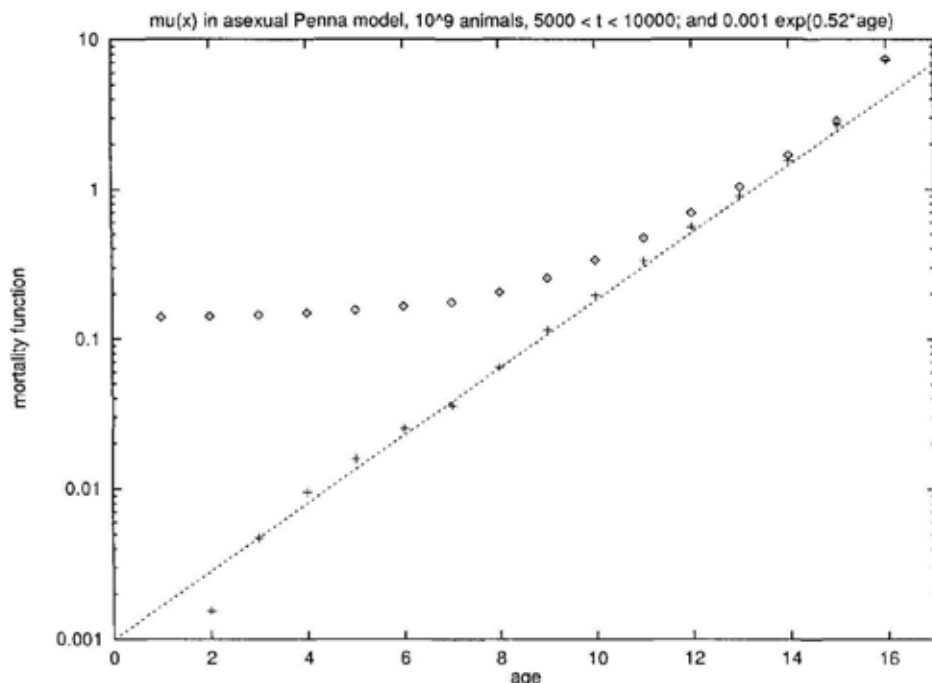


Figure 5. Mortality function  $\mu$  for the standard asexual Penna model, with (upper data) and without (lower data) the deaths from the Verhulst factor.

Figure 5 from <sup>35</sup> shows the resulting mortalities, with and without the deaths from starvation or lack of space. The purely genetic deaths follow nicely the exponential Gompertz law, except for the youngest and oldest ages. The limit of exactly three mutations killing can be softened <sup>38</sup> to give slight downward deviations from the Gompertz law as in Swedish mortalities, Fig. 1, or a mortality plateau as claimed in <sup>10</sup>. It can also be abolished in favour of Verhulst factors depending on the number of active mutations; then a mortality maximum even more pronounced than for flies <sup>8,9</sup> is obtained in Fig. 6 from <sup>39</sup>. Simulations of biologists, in contrast, could not yet get such mortality maximum <sup>40</sup>. Pacific salmon, Northern cod, and Alaskan wolves were simulated successfully long ago <sup>35</sup>, and Lyapunov exponents <sup>45</sup>, Brazilian lobsters <sup>42</sup>, child mortality <sup>46,47</sup>, prey-predator relations on lattice <sup>43</sup>, and speciation more recently <sup>41,44</sup>. Particularly relevant for our section 2 are the Penna model simulations of the demographic changes in the 20th century <sup>48</sup>. The mortalities do not change much <sup>49</sup> if the genome may contain the same gene in several copies called "paralogs" <sup>21</sup>.

This section ends with a technical warning: If the population is kept constant artificially, as is tradition in theoretical biology, instead of being allowed to fluctuate as in nature, then the results are only qualitatively, not quantitatively, the same <sup>50</sup>.

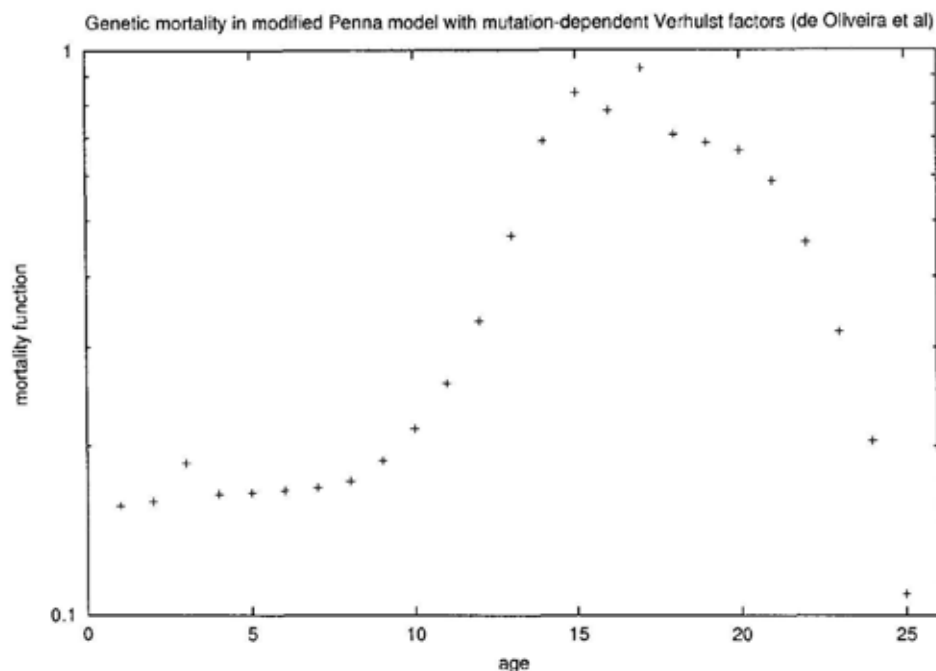


Figure 6. Semilogarithmic plot of mortality function in a modified Penna model

## 5 Sex

Sexual reproduction was introduced into the Penna model long ago<sup>35</sup>, for the Partridge-Barton type<sup>34</sup> and the Stauffer model<sup>26</sup> only recently. Even bacteria exchange genome via "parasex"<sup>51</sup>, and computer simulations with the Penna model showed this parasex to give fitter individuals than pure asexual cloning<sup>52</sup>. These simulations included ageing of bacteria, as found later experimentally<sup>53</sup>. Less clear is the need for males in species with two sets of chromosomes, from father and mother<sup>36</sup>. Only with some effort<sup>55</sup> could feeding the males be justified in the Penna model; and no simulation yet showed hermaphroditism to be by far the fittest way of reproduction. On the other hand, sexual reproduction is clearly preferable as a protection against parasites or environmental catastrophes<sup>36</sup>. So, the sex wars can continue<sup>54</sup>.

Menopause or its analog is the cessation of female reproductive power at middle age. In spite of widespread prejudice, it is not restricted to humans (and pilot whales) but even occurs in some flies<sup>56</sup>. Computer simulations showed, without any specific human assumptions like tradition of knowledge, that menopause can emerge automatically<sup>57,26</sup>, provided the risk for the mother of giving birth increases with increasing age and/or the child depends on the mother for its initial survival. Are men needed for survival of the children? Only indirectly<sup>35</sup>: without them women would follow Pacific salmon and die rapidly after giving birth; the lack of

a male analog for a sharp menopause makes males useful for producing children even at older age, thus prevented evolution to kill females after their cessation of reproduction.

## 6 Conclusion

Computer simulation of mutation accumulation models has advanced a lot in one decade and has applications like retirement rules. Particularly important seem the menopause explanations<sup>57,26</sup> showing that such effects are not restricted to humans. Simulations of alternative theories of biological ageing<sup>19</sup> are mostly lacking.

## Acknowledgements

I thank N. Jan, S. Moss e Oliveira, P.M.C. de Oliveira, T.J.P.Penna, A.T. Bernardes, S. Cebrat. J.S. Sá Martins, A.O. Sousa and many others for fruitful collaboration over many years.

## References

1. A. R. Thatcher, V. Kannisto and J. W. Vaupel, *The Force of Mortality at Ages 80 to 120* (Odense University Press, Odense, 1998).
2. J.W. Vaupel et al., *Science* **280**, 855 (1998).
3. B.L. Strehler, A.S. Mildvan, *Science* **132**, 142 (1960).
4. L.A. Gavrilov, N.S. Gavrilova, *J. Theor. Biology* **213**, 527 (2001).
5. M. Ya. Azbel, *Proc. Roy. Soc. B* **263**, 1449 (1996).
6. D. Stauffer, *Exp. Gerontology* **37**, 1131 (2002); J.S. Sá Martins, D. Stauffer, Ingenierias (Univ. Nuevo Leon, Mexico), in press.
7. M. Ya. Azbel, *Phys. Rev. E* **66**, 016107 (2002).
8. J. R. Carey, P. Liedo, D. Orozco, and J. W. Vaupel, *Science* **258**, 457 (1992).
9. J.W. Curtsinger, H.H. Fukui, D.R. Townsend, J.W. Vaupel, *Science* **258**, 461 (1992).
10. J.-M. Robine, J.W. Vaupel, *Exp. Gerontology* **36**, 915 (2001).
11. D. Stauffer, in *Annual Reviews of Computational Physics*, vol. VIII, p. 329 (World Scientific, Singapore, 2000).
12. D. Stauffer, in *Biological Evolution and Statistical Physics*, edited by M. Lässig and A. Valleriani, p.258 (Springer, Berlin-Heidelberg, 2002).
13. T.T. Perls, *Sci. Amer.* **272**, 50 (Jan. 1995).
14. M. Heumann and M. Hötzel, *J. Stat. Phys.* **79**, 483 (1995).
15. W. Mair, P. Goymer, S.D. Pletcher and L. Partridge, *Science* **301**, 1617 (2003).
16. K.T. Howitz et al., *Nature* **425**, 191 (2003).
17. A. Weismann, *Über die Dauer des Lebens* (Gustav Fischer, Jena, 1882).
18. P.B. Medawar, *Modern Q.* **1**, 30 (1946).
19. A. Aviv, D. Levy and M. Mangel, *Mech. Ageing and Development* **124**, 929 (2003); Z.Tan, *Exp. Gerontology* **36**, 2001 (89).
20. S.D. Pletcher, D. Neuhauser, *Int. J. Mod. Phys.C* **11**, 525 (2000), Z.F. Huang, D. Stauffer, *Theory in Biosciences* **120**, 21 (2001).

21. S. Cebrat, D. Stauffer, J. Appl. Genetics (Poznan) **43**, 391 (2002).
22. L. Partridge and N.H. Barton, Nature **362**, 305 (1993).
23. T. J. P. Penna, J. Stat. Phys. **78**, 1629 (1995).
24. H. Meyer-Ortmanns, Int. J. Mod. Phys. C **12**, 319 (2001).
25. D. Stauffer and J.P. Radomski, Exp. Gerontol. **37**, 175 (2001); T. Shimada, Int. J. Mod. Phys. C **12**, 1207 (2001).
26. A.O. Sousa, Physica A **326** **233**, 2003 (;) A.O. Sousa, S. Moss de Oliveira, D. Stauffer, Int. J. Mod. Phys.C **12**, 1477 (2001).
27. D. Makowiec, D. Stauffer and M. Zielinski, Int. J. Mod. Phys. C **12**, 1067 (2001).
28. D. Chowdhury and D. Stauffer, Phys. Rev. E, in press and Physica A, in press.
29. A.O. Sousa, S. Moss de Oliveira, Physica A **294**, 431 (2001).
30. T.S. Ray, J. Stat. Phys. **74**, 929 (1994).
31. N. Jan, J. Stat. Phys. **77**, 915 (1994).
32. S. Vollmar and S. Dasgupta, J. Physique I **4**, 817 (1994); S. Dasgupta, J. Physique I **4**, 1563 (1994) and Physica Scripta T **106**, 19 (2003).
33. N.G.F. Medeiros, R.N. Onody, Phys. Rev. E **64**, 041915 (2001).
34. A.O. Sousa, Theory in Biosciences, in press.
35. S. Moss de Oliveira, P.M.C. de Oliveira and D. Stauffer, *Evolution, Money, War and Computers* (Teubner, Leipzig, 1999).
36. D. Stauffer, P.M.C. de Oliveira, S. Moss de Oliveira, T.J.P. Penna, J.S. Sa Martins, An. Acad. Bras. Ci. **73**, 15 (2001) = cond-mat/0011524.
37. A.O. Sousa, S. Moss de Oliveira, J.S.S. Martins, Phys. Rev. E **67**, 032903 (2003).
38. J.B. Coe, Y. Mao, M.E. Cates, Phys. Rev. Lett. **89**, 288103 (2002).
39. S. Moss de Oliveira, D. Stauffer, P.M.C de Oliveira, J.S. Sá Martins, Physica A, in press.
40. L.D. Mueller et al., Exp. Gerontology **38**, 373 (2003); B.Charlesworth., J. Theor. Biol. **210**, 47 (2001).
41. A. Łaszkiwicz, Sz. Szymczak, S. Cebrat, Int. J. Mod. Phys. C **14**, No. 6 (2003).
42. T.J.P. Penna, A. Racco, A.O. Sousa, Physica A **295**, 31 (2001).
43. M.F. He, J. Lin, H. Jiang and X. Liu, Physica A **312**, 243 (2002).
44. K. Luz-Burgoa, S. Moss de Oliveira, J.S. Sá Martins, D. Stauffer, A.O. Sousa, Braz. J. Phys. **33**, in press (2003).
45. A. Castro-e-Silva, A.T. Bernardes, Physica A **301**, 63 (2001).
46. A. Łaszkiwicz, Sz. Szymczak,, A. Kurdziel, S. Cebrat, Int. J. Mod. Phys.C **13**, 97 (2002).
47. M.S. Magdon-Maksymowicz, M. Sitarz, M. Bubak, A.Z. Maksymowicz, J. Szewczyk, Computer Phys. Comm. **147**, 621 (2002).
48. A. Łaszkiwicz, Sz. Szymczak, S. Cebrat, Int. J. Mod. Phys. **14**, No. 10 (2003), and Theory in Biosciences, in press = condmat/0305277.
49. P. Alle, masters thesis, Cologne University, 2003.
50. S. Moss de Oliveira, P.M.C. de Oliveira, and J.S. Sá Martins, Int. J. Mod. Phys. C **15**, No. 2 (2004) = cond-mat/0308600.

51. M. Kohiyama, S. Hiraga, I. Matic and M. Radman, *Science* **301**, 802 (2003).
52. S. Moss de Oliveira, P.M.C. de Oliveira, D. Stauffer, *Physica A* **322**, 521 (2003).
53. M. Ackermann, S.C. Stearns, U. Jenal, *Science* **300**, 1920 (2003).
54. P.M.C. de Oliveira, S. Moss de Oliveira, *Int. J. Mod. Phys. C* **14**, 241 (2003).
55. J.S. Sá Martins, D Stauffer, *Physica A* **294**, 191 (2001).
56. V.N. Novoseltsev et al., *Mech. Ageing Dev.* **124**, 605 (2003), Fig. 5.
57. S. Moss de Oliveira, A.T. Bernardes and J.S. Sá Martins, *Eur. Phys. J. B* **7**, 510 (1999).

# FITTING CURVES BY FRACTAL INTERPOLATION: AN APPLICATION TO THE QUANTIFICATION OF COGNITIVE BRAIN PROCESSES

M.A. NAVASCUES

*Departamento de Matemática Aplicada, Centro Politécnico Superior de Ingenieros,  
Universidad de Zaragoza, C/ María de Luna, 3. 50018 Zaragoza, Spain  
E-mail: manavas@unizar.es*

M.V. SEBASTIAN

*Departamento de Matemáticas, Universidad de Zaragoza, Edificio de Matemáticas,  
Ciudad Universitaria s/n, 50009 Zaragoza, Spain  
E-mail: msebasti@unizar.es*

Fractal interpolants of Barnsley are defined for any continuous function defined on a real compact interval. The uniform distance between the function and its approximant is bounded in terms of the vertical scale factors. As a general result, the density of the affine fractal interpolation functions of Barnsley in the space of continuous functions in a compact interval is proved. A method of data fitting by means of fractal interpolation functions is proposed. The procedure is applied to the quantification of cognitive brain processes. In particular, the increase in the complexity of the electroencephalographic signal produced by the execution of a test of visual attention is studied. The experiment was performed on two types of children: a healthy control group and a set of children diagnosed with an attention deficit disorder.

## 1 Introduction

Barnsley<sup>1</sup> was a pioneer in the use of fractal functions to interpolate a set of data. His method constitutes an advance in the techniques of approximation in the sense that the interpolants used are not necessarily differentiable and show the rough aspect of the real-world signals. Besides, by means of that procedure any other interpolation, polynomial spline for instance, can be generalized<sup>2</sup>.

We use that methodology to define fractal functions approximating any continuous function by means of a suitable iterated function system. The uniform distance between the function and its fractal analogue is also bounded. As a particular case, the affine fractal interpolation functions are the generalization of the polygonal (piecewise linear) functions. As a consequence of the inequalities obtained, the density of the affine fractal functions of Barnsley in the space of continuous functions in a compact interval is deduced.

As an application, a method to fit real data is proposed here. The procedure is applied to the quantification of cognitive brain processes. In particular, the increase in the complexity of the electroencephalographic signals during the testing of visual attention is studied. The experiment was performed on two types of children: a healthy control group and a set of children diagnosed with an Attention Deficit with Hyperactivity Disorder (ADHD).



## 2 Fractal Interpolation of a Continuous Function in a Compact Real Interval

### 2.1 Generalization of continuous functions

Let  $\Delta : t_0 < t_1 < \dots < t_N$  be a partition of the closed interval  $I = [t_0, t_N]$ . Let a set of data points  $\{(t_n, x_n) \in I \times R : n = 0, 1, 2, \dots, N\}$  be given. Set  $I_n = [t_{n-1}, t_n]$  and let  $L_n : I \rightarrow I_n$ ,  $n \in \{1, 2, \dots, N\}$  be contractive homeomorphisms such that:

$$L_n(t_0) = t_{n-1}, \quad L_n(t_N) = t_n \quad (1)$$

$$|L_n(c_1) - L_n(c_2)| \leq l |c_1 - c_2| \quad \forall c_1, c_2 \in I \quad (2)$$

for some  $0 \leq l < 1$ .

Let  $-1 < \alpha_n < 1$ ;  $n = 1, 2, \dots, N$ ,  $F = I \times [c, d]$  for some  $-\infty < c < d < +\infty$  and  $N$  continuous mappings,  $F_n : F \rightarrow R$  be given satisfying:

$$F_n(t_0, x_0) = x_{n-1}, \quad F_n(t_N, x_N) = x_n, \quad n = 1, 2, \dots, N \quad (3)$$

$$|F_n(t, x) - F_n(t, y)| \leq \alpha_n |x - y|, \quad t \in I, \quad x, y \in R \quad (4)$$

Now define functions  $w_n(t, x) = (L_n(t), F_n(t, x))$ ,  $\forall n = 1, 2, \dots, N$ .

**Theorem** (Barnsley<sup>1</sup>): The iterated function system (IFS)<sup>3</sup>  $\{F, w_n : n = 1, 2, \dots, N\}$  defined above admits a unique attractor  $G$ .  $G$  is the graph of a continuous function  $f : I \rightarrow R$  which obeys  $f(t_n) = x_n$  for  $n = 0, 1, 2, \dots, N$ .

The previous function is called a fractal interpolation function (FIF) corresponding to  $\{(L_n(t), F_n(t, x))\}_{n=1}^N$ .

Let  $\mathcal{G}$  be the set of continuous functions  $f : [t_0, t_N] \rightarrow [c, d]$  such that  $f(t_0) = x_0$ ;  $f(t_N) = x_N$ .  $\mathcal{G}$  is a complete metric space respect to the uniform norm. Define a mapping  $T : \mathcal{G} \rightarrow \mathcal{G}$  by:

$$(Tf)(t) = F_n(L_n^{-1}(t), f \circ L_n^{-1}(t)) \quad \forall t \in [t_{n-1}, t_n], \quad n = 1, 2, \dots, N \quad (5)$$

$T$  is a contraction mapping on the metric space  $(\mathcal{G}, \|\cdot\|_\infty)$ :

$$\|Tf - Tg\|_\infty \leq |\alpha|_\infty \|f - g\|_\infty \quad (6)$$

where  $|\alpha|_\infty = \max \{|\alpha_n|; n = 1, 2, \dots, N\}$ . Since  $|\alpha|_\infty < 1$ ,  $T$  possesses a unique fixed point on  $\mathcal{G}$ , that is to say, there is  $f \in \mathcal{G}$  such that  $(Tf)(t) = f(t) \quad \forall t \in [t_0, t_N]$ . This function is the FIF corresponding to  $w_n$  and it is the unique  $f \in \mathcal{G}$  satisfying the functional equation<sup>1</sup>:

$$f(t) = F_n(L_n^{-1}(t), f \circ L_n^{-1}(t)), \quad n = 1, 2, \dots, N, \quad t \in I_n = [t_{n-1}, t_n] \quad (7)$$

The most widely studied fractal interpolation functions so far are defined by the IFS

$$\begin{cases} L_n(t) = a_n t + b_n \\ F_n(t, x) = \alpha_n x + q_n(t) \end{cases} \quad (8)$$

with

$$a_n = \frac{(t_n - t_{n-1})}{(t_N - t_0)} \quad \text{and} \quad b_n = \frac{(t_N t_{n-1} - t_0 t_n)}{(t_N - t_0)} \quad (9)$$

$\alpha_n$  is called a vertical scaling factor of the transformation  $w_n$  and  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$  is the scale vector of the IFS. If  $q_n(t)$  is a line, the FIF is termed affine.

Barnsley proposes, in the reference [1], the generalization of a continuous function  $g$  by means of a fractal interpolation defined by the IFS (8) with

$$q_n(t) = g \circ L_n(t) - \alpha_n b(t), \quad (10)$$

where  $b$  is continuous and such that  $b(t_0) = x_0$  and  $b(t_N) = x_N$ . It is easy to check that, in this case, (8) verifies the hypotheses of Barnsley's theorem.

**Definition 1:** Let  $g \in C(I)$ ,  $\Delta$ ,  $b$  and  $\alpha$  as in the previous paragraphs. The FIF  $g_{\Delta b}^\alpha$  or simply  $g^\alpha$  defined by (8), (9) and (10) is termed  $\alpha$ -fractal function of  $g$  with respect to  $\Delta$  and  $b$ .

**Theorem 1:** The  $\alpha$ -fractal function  $g^\alpha$  of  $g$  with respect to  $\Delta$  and  $b$  satisfies the inequality

$$\|g^\alpha - g\|_\infty \leq \frac{|\alpha|_\infty}{1 - |\alpha|_\infty} (\|g - b\|_\infty) \quad (11)$$

with  $|\alpha|_\infty = \max_{1 \leq n \leq N} \{|\alpha_n|\}$ . Besides,  $g^\alpha$  interpolates  $g$ , that is to say,

$$g^\alpha(t_n) = g(t_n) \quad \forall n = 0, 1, \dots, N \quad (12)$$

**Proof:** Let  $g \in C(I)$  be a continuous function and consider an IFS as in (8), (9) and  $q_n(t) = g \circ L_n(t) - \alpha_n b(t)$ ,  $\forall n = 1, 2, \dots, N$ .

For  $\mathcal{G} = \{f : [a, b] \rightarrow [c, d] \text{ cont}; f(a) = x_0, f(b) = x_N\}$  define  $T_\alpha : \mathcal{G} \rightarrow \mathcal{G}$

$$T_\alpha f(t) = F_n(L_n^{-1}(t), f \circ L_n^{-1}(t)) \quad \forall t \in I_n \quad (13)$$

That is to say,

$$T_\alpha f(t) = g(t) + \alpha_n(f - b) \circ L_n^{-1}(t) \quad \forall t \in I_n \quad (14)$$

By the theorem of Barnsley,  $T_\alpha$  admits a unique fixed point in  $\mathcal{G}$  continuous on  $I$  ( $g^\alpha$ ).  $g^\alpha$  is defined by the equation:

$$g^\alpha(t) = g(t) + \alpha_n(g^\alpha - b) \circ L_n^{-1}(t) \quad \forall t \in I_n \quad (15)$$

From (15):

$$\|g^\alpha - g\|_\infty \leq |\alpha|_\infty (\|g^\alpha - b\|_\infty) \leq |\alpha|_\infty (\|g^\alpha - g\|_\infty + \|g - b\|_\infty) \quad (16)$$

and the inequality (11) is proved.

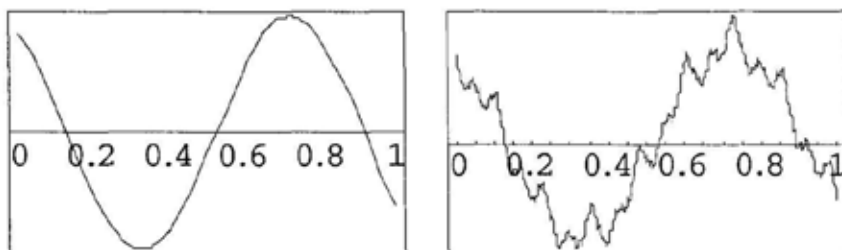


Figure 1. The left figure represents the graph of the function  $g(t) = \sin(8(t + \pi/3))$ . The right figure shows the corresponding  $\alpha$ -function, with respect to  $\Delta : 0 < 1/8 < 2/8 < \dots < 1$ ,  $b$  a line and  $\alpha_n = 0.2 \quad \forall n = 1, \dots, 8$ .

$g^\alpha$  passes through the points  $(t_n, x_n)$  as, by (1) and (3):

$$g^\alpha(t_n) = F_n(L_n^{-1}(t_n), g^\alpha \circ L_n^{-1}(t_n)) = F_n(t_N, g^\alpha(t_N)) = F_n(t_N, x_N) = x_n \quad (17)$$

◇

## 2.2 Affine fractal interpolation functions

If  $g$  is piecewise linear in the intervals  $I_n = [t_{n-1}, t_n]$  for  $n = 1, 2, \dots, N$  and  $b$  is the line between  $(t_0, x_0)$  and  $(t_N, x_N)$ ,  $q_n$  is a first degree polynomial and  $g^\alpha$  is the affine fractal interpolation respect to the data  $\{(t_n, x_n = g(t_n)), n = 0, 1, \dots, N\}$  with scale vector  $\alpha$ . In this case, by (11) one has

$$\|g^\alpha - g\|_\infty \leq \frac{2|\alpha|_\infty}{1 - |\alpha|_\infty} \max_{0 \leq n \leq N} \{|x_n|\} \quad (18)$$

**Theorem 2<sup>4</sup>:** Let  $\{a, b, t_1, t_2, t_3, \dots\}$  be a dense sequence in  $I = [a, b]$ . For each  $n \geq 1$  let  $\phi_n$  denote the polygonal (piecewise linear) function which agrees with  $f$  at nodes  $a, b, t_1, \dots, t_n$  (conveniently ordered). Then  $\phi_n \rightarrow f$  as  $n \rightarrow \infty$  for all  $f \in \mathcal{C}(I)$ .

**Theorem 3:** The affine fractal interpolation functions are dense in  $\mathcal{C}[a, b]$ .

Proof: Let  $f$  be a continuous function in  $\mathcal{C}[a, b]$  and  $\{a, b, t_1, t_2, t_3, \dots\}$  dense in  $[a, b]$ . Given  $\varepsilon > 0$ , consider  $\varepsilon/2 > 0$ , by the theorem 2 there exists  $n \in \mathbb{N}$  such that the polygonal function  $\phi_n$  which agrees with  $f$  at  $a, b, t_1, \dots, t_n$  verifies:

$$|f(t) - \phi_n(t)| < \varepsilon/2 \quad \forall t \in [a, b] \quad (19)$$

Choose  $\alpha \in \mathbb{R}^{n+1}$  such that  $\alpha \neq 0$  and

$$\frac{2|\alpha|_\infty}{1 - |\alpha|_\infty} \max\{|f(t)|; t = a, b, t_1, \dots, t_n\} < \varepsilon/2 \quad (20)$$

Let  $\phi_n^\alpha(t)$  the  $\alpha$ -fractal function of  $\phi_n$  with respect to  $\{a, b, t_1, \dots, t_n\}$  and the line  $b$  passing through  $f(a), f(b)$ .  $\phi_n^\alpha(t)$  is an affine fractal interpolation function as seen before and by (18), (19) and (20)

$$|f(t) - \phi_n^\alpha(t)| \leq |f(t) - \phi_n(t)| + |\phi_n(t) - \phi_n^\alpha(t)| < \varepsilon \quad (21)$$

◇

### 3 Fit of sampled data by fractal interpolation

The description of a procedure to fit a set of real data  $\{(t_i^*, y_i^*), i = 0, 1, \dots, M\}$  by means of fractal interpolation functions is developed here. This function is constructed as perturbation of an interpolant  $g$  of a subsample of the data.

Let  $P = \{(t_n, x_n), n = 0, 1, \dots, N\}$  be a subset of the data with  $(t_0, x_0) = (t_0^*, y_0^*)$  and  $(t_N, x_N) = (t_M^*, y_M^*)$  and let  $g$  be an interpolation function passing through  $P$ . We consider here an IFS (8) with (9) and  $q_n(t) = g \circ L_n(t) - \alpha_n b(t)$ ,  $b$  continuous and such that  $b(t_0) = x_0$ ,  $b(t_N) = x_N$  and  $g^\alpha$  the corresponding FIF.

Let  $\{(\bar{t}_j^n, \bar{x}_j^n), j = 1, 2, \dots, m^{(n)}\}$  the intermediate points in  $I_n = [t_{n-1}, t_n]$  not included in  $P$ :

$$t_{n-1} \leq \bar{t}_j^n \leq t_n \quad \forall j = 1, \dots, m^{(n)} \quad (22)$$

According to the equality (15), with the condition  $g^\alpha(\bar{t}_j^n) = \bar{x}_j^n$ :

$$\bar{x}_j^n = g(\bar{t}_j^n) + \alpha_n(g^\alpha - b) \circ L_n^{-1}(\bar{t}_j^n) \quad (23)$$

and

$$\bar{x}_j^n \simeq g(\bar{t}_j^n) + \alpha_n(g - b) \circ L_n^{-1}(\bar{t}_j^n) \quad (24)$$

Choose  $\alpha_n$  so that the sum of the square residuals be minimum:

$$\min E(\alpha_n) = \sum_{j=1}^{m^{(n)}} (g(\bar{t}_j^n) - \bar{x}_j^n + \alpha_n(g - b) \circ L_n^{-1}(\bar{t}_j^n))^2 \quad (25)$$

The equation  $E'(\alpha_n) = 0$  gives:

$$\alpha_n = - \frac{\sum_{j=1}^{m^{(n)}} (g(\bar{t}_j^n) - \bar{x}_j^n)(g(L_n^{-1}(\bar{t}_j^n)) - b(L_n^{-1}(\bar{t}_j^n)))}{\sum_{j=1}^{m^{(n)}} (g(L_n^{-1}(\bar{t}_j^n)) - b(L_n^{-1}(\bar{t}_j^n)))^2} \quad (26)$$

Denoting

$$\varepsilon = (g(\bar{t}_1^n) - \bar{x}_1^n, \dots, g(\bar{t}_{m^{(n)}}^n) - \bar{x}_{m^{(n)}}^n) \quad (27)$$

$$\varepsilon_b = (g(L_n^{-1}(\bar{t}_1^n)) - b(L_n^{-1}(\bar{t}_1^n)), \dots, g(L_n^{-1}(\bar{t}_{m^{(n)}}^n)) - b(L_n^{-1}(\bar{t}_{m^{(n)}}^n))) \quad (28)$$

then

$$|\alpha_n| = \frac{|\varepsilon \cdot \varepsilon_b|}{|\varepsilon_b|^2} \leq \frac{|\varepsilon|}{|\varepsilon_b|} \quad (29)$$

If the interpolant  $g$  converges to the original function when the diameter of the partition tends to 0, then  $(g(\bar{t}_j) - \bar{x}_j) \rightarrow 0$  and  $\varepsilon \rightarrow 0$ . Therefore,  $\alpha_n$  can be chosen so that  $|\alpha_n| < 1$ ,  $\forall n = 1, 2, \dots, N$ . The function  $g$  and the scale vector  $\alpha$  determine the fitting curve  $g^\alpha$ .

### 3.1 Error bound of the fit

Let  $h$  be the original function corresponding to the data and  $g$  the interpolant used, let  $g^\alpha$  be the fitting function defined here. Then:

$$\|h - g^\alpha\|_\infty \leq \|h - g\|_\infty + \|g - g^\alpha\|_\infty \leq K_g + \frac{|\alpha|_\infty}{1 - |\alpha|_\infty} (\|g - b\|_\infty) \quad (30)$$

with  $K_g$  an upper bound of the interpolation error respect to  $g$ .

### 3.2 A particular case

If  $m^{(n)} = 1$  then, from (23):

$$\alpha_n = \frac{g(\bar{t}_1^n) - \bar{x}_1^n}{b \circ L_n^{-1}(\bar{t}_1^n) - g^\alpha \circ L_n^{-1}(\bar{t}_1^n)} \quad (31)$$

If  $g$  is piecewise linear,  $N$  is even,  $b$  is the straight line passing through  $(t_0, x_0)$ ,  $(t_N, x_N)$  and  $\bar{t}_1^n$  is chosen as

$$\bar{t}_1^n = \frac{t_{n-1} + t_n}{2}$$

then  $L_n^{-1}(\bar{t}_1^n) = (t_0 + t_N)/2 = t_{N/2}$ ,  $g^\alpha \circ L_n^{-1}(\bar{t}_1^n) = x_{N/2}$  and the formula of Strahle<sup>5</sup> is obtained

$$\alpha_n = \frac{\frac{x_{n-1} + x_n}{2} - \bar{x}_1^n}{\frac{x_0 + x_N}{2} - x_{N/2}} \quad (32)$$

## 4 Application to the quantification of cognitive brain processes

### 4.1 Methods

The computation of the correlation dimension of the EEG recordings has been widely treated and used by the neurophysiologist community (see, for instance, the references Babloyantz et al.<sup>6</sup>, Stam et al.<sup>7</sup> and Nan and Jinghua<sup>8</sup>). However, some controversy is still alive because of the algorithmic and modelling problems. We propose here an alternative to the computation of the correlation dimension of the attractor reconstructed by means of the method of Takens. The dimension of the graph of the EEG, as geometric object of  $R^2$ , can be seen as a measure of the complexity of the bioelectric brain signals. The first step is to reconstruct the signal by means of fractal interpolation functions, computing the parameters of the IFS associated with the data according to the fit proposed in the previous paragraph.

Following a theorem of Barnsley<sup>3</sup> (see also Besicovitch and Ursell<sup>9</sup>) the fractal dimension  $D$  of the graph of an affine FIF verifies the equation

$$\sum_{n=1}^N |\alpha_n| a_n^{D-1} = 1 \quad (33)$$

with  $\alpha_n$  and  $a_n$  being the coefficients defined in the paragraph 2.1 ((8),(9)). If the nodes are equidistant,  $a_n = 1/N$  and

$$D = 1 + \frac{\log(\sum_{n=1}^N |\alpha_n|)}{\log N} \quad (34)$$

This formula for the dimension is valid in the case  $\sum_{n=1}^N |\alpha_n| > 1$ . Otherwise, the fractal dimension is one<sup>3</sup>. This parameter lies between 1 and 2. Its computation is simple, the use of delay variables is not necessary and there are not problems of convergence or insufficient number of points.

## 4.2 Materials

In the present work, those procedures are applied to the study of the EEG recordings of two samples of children: a healthy control group and a set diagnosed with an Attention Deficit with Hyperactivity Disorder (ADHD). The clinical manifestations of the ADHD are characterised by a lack of attention, impulsive cognitive and behaviour styles and by an excessive motor activity. Its incidence is estimated between 3 and 5 % of the school population and one or two children with deficient attention per classroom during the first school years could be observed. By a mere visual inspection of the EEG, no difference was observed in the patient group.

The children belonging to the control group were selected randomly by the teachers and belong to the same school groups as the children with ADHD. 19 children diagnosed with ADHD were chosen, with an average age of 9.3 and a standard deviation of 1.5. The sample was compared with a control group of 13 children with similar age (9.2) and standard deviation (1.3).

For every subject, the following signals were recorded: (i) an EEG at rest with closed eyes, (ii) an EEG during the execution of a test consisting in the recognition of a face different from the others, in series of three.

Six locations of the cortical surface were analyzed, following the 10-20 International System of Jasper: F3, F4, P3, P4, O1, O2. The recording of the signal was performed by an electroencephalograph Grass, connected to the program RHYTHM, version 5. The equipment included filters of 0.18 Hz for low frequencies and 35 Hz for high frequencies. The sensitivity is 7 microvolts per millimeter. The sampling frequency was 128 Hz. Every segment chosen for spectral analysis had a length of 4 seconds, but the results are not here due to space restrictions. A segment of 30 seconds was analyzed within the second minute.

The fractal dimension of the EEG was obtained by the method proposed in the section 3, considering an affine FIF and two inner points between every pair of points used for the fractal interpolation ( $m^{(n)} = 2$ ). To compare the EEG at rest with the EEG recorded while the execution of the described exercise, the Wilcoxon

Table 1. Average values of the fractal dimension for each group and EEG.

	Control		Deficient Attention	
	Rest	Test	Rest	Test
F3	1.7240	1.7311	1.7043	1.7226
F4	1.7120	1.7413	1.6736	1.7334
P3	1.7377	1.7449	1.6931	1.7410
P4	1.6987	1.7616	1.6868	1.7218
O1	1.7024	1.7509	1.7166	1.7843
O2	1.6929	1.7940	1.7197	1.7841

sign hierarchized test was used.

### 4.3 Results and discussion

The table 1 shows the average values of the fractal dimension for each group, EEG and channel. This parameter increases on the whole cortical surface during the execution of the visual test, but the difference is only significant in some locations.

Comparing the data obtained in the computation of the fractal dimension of the EEG during the face recognition test respect to the rest EEG, differences were found, in the control group, in O2 with a significance level of 0.01. In the group of children with ADHD the differences occurred in F4, O1 and O2 at level 0.05 and O1 and O2 at level 0.01. These variations show the activation of the occipital zone (primary visual area) in the achievement of tasks of visual attention, as well as the need of the children with ADHD to activate more cortical networks to perform the same test (fig 2). These results are coherent with the findings of our group<sup>10,11</sup> by means of other techniques (spectral,  $D_2$ , Hjorth).

The results described aim at a lower dimensionality in the rest EEG. This fact coincides with the studies of Graf and Elbert<sup>12</sup>, Nan and Jinghua<sup>8</sup> and Pritchard and Duke<sup>13</sup> for different pathologies and brain processes. The children with deficient attention show also a degree and extension of the cortical activity higher in the execution of the visual task. As a consequence, the children with ADHD would need to activate a higher number of cortical networks in the processing of the same information, and this fact is expressed in the changes in the brain electrogenesis. To analyse conveniently some information they would need a higher amount of energy than the other children.

The non-linear measure most widely known in electroencephalography is the correlation dimension. It has been used to evaluate the bioelectric and cognitive activities. Nan and Jinghua<sup>8</sup> describe an increasing of the correlation dimension by mental tasks of arithmetic character, with respect to rest activity. Gregson et al.<sup>14</sup> find also an increase in the dimension during experiments of visual attention.

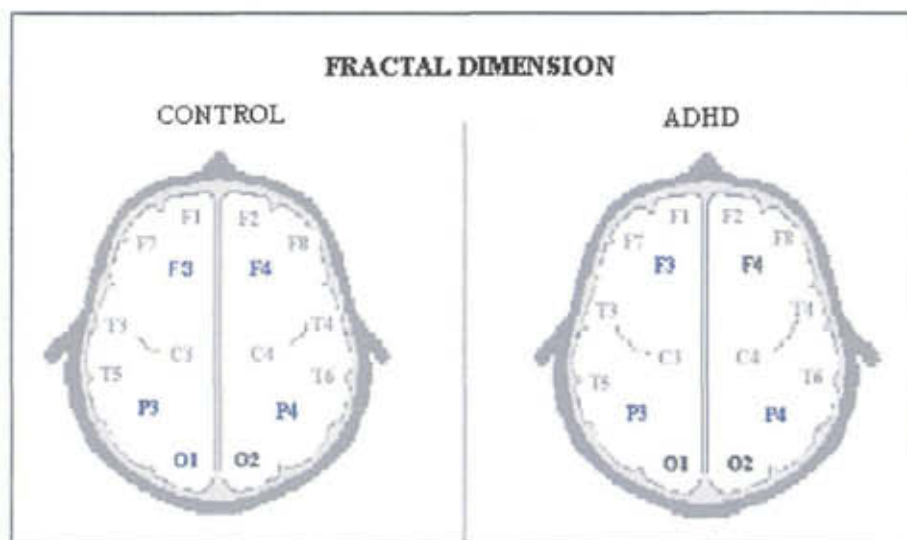


Figure 2. In blue, electrodes where there was not significant difference between rest and visual task EEGs, in black those displaying differences, in the rest the study was not performed.

The FIF dimensions allow to show the complexity increase in the stimulated zones with a lower computational and algorithmic cost constituting this way an efficient measure of the neural bioelectric activity.

#### 4.4 Other techniques

The computation of the fractal dimensions of reconstructed attractors is widely known in the neurophysiological literature. A collection of procedures can be reviewed in the book of Parker & Chua<sup>15</sup>. The capacity, the information dimension or the correlation dimension are members of an infinite family called Renyi<sup>16</sup> dimensions. Their calculation is usually based on the method of embedding (Takens<sup>17</sup>). From a single sampled signal, a whole trajectory in a higher-dimensional space is reconstructed, considering as coordinates the delays of the recording. This path allows to estimate fractal dimensions by means of some numerical methods, for instance the algorithm of Grassberger & Procaccia<sup>18</sup>.

However, the measures of experimental data are obstructed by the presence of noise, either inherent in the system or provided by the instrumental devices. Those factors obscure the structure of the possible underlying attractor. This fact generates a number of algorithmic problems some of which we briefly summarize here (see also Mayer-Kress<sup>19</sup>):

- The number of points necessary to provide a reliable estimation of the parameter can be very large. For instance, in an article of Smith<sup>20</sup> a lower bound of



this quantity is given by the inequality  $N_{min} \geq 42^M$ , with  $M$  the greatest integer lower than the dimension. In the case of EEG signals, this estimation supposes to handle dozens of millions of points, which usually are not provided by the conventional recordings. The alternative proposed in the same paper is to compute a local scaling around some points of the flow, that supposes to deal with multivariant measures.

- The need to reach a convergence value for the dimension with respect to the number of delay variables is not fulfilled in general, as communicated in many papers about the subject (Basar et al.<sup>21</sup>, Lopes da Silva et al.<sup>22</sup>). This fact is related to the choice of an optimum embedding (see, for instance, Ding et al.<sup>23</sup>). Even in the case of existence of limit, the computational cost can be unfeasible for a standard Department of Neurophysiology<sup>15</sup>.

- There are no standard criteria to choose other algorithmic parameters as, for instance, the length of lag interval between coordinates, the sampling frequency of the signal, the scale range to perform the linear fit (Ellner<sup>24</sup>) or the filters to use (Basar et al.<sup>21</sup>). As a consequence, the differences in the reported EEG dimensions for the same process and/or pathology are excessively large (Basar et al.<sup>21</sup>, Holzfuss & Mayer-Kress<sup>25</sup>).

- To avoid these difficulties several complementary methods have been proposed. One of them is the singular spectrum analysis (Broomhead & King<sup>26</sup>) which pursues the calculation of uncorrelated variables and an estimation of the embedding dimension. Additionally, the process performs an adaptive moving-average filtering associated with the dominant oscillations of the system. Another method is the "projection pursuit", that uses low-dimensional projections to form an estimate of the probability density (Friedman et al.<sup>27</sup>). However, at the moment these procedures are not conclusive (see for instance Mees et al.<sup>28</sup>).

We propose here an alternative methodology: the calculation of the fractal dimension of the EEG as the attractor of an explicit IFS, instead of the projection of a high-dimensional limit set. Relevant contributions in this area are given by Mandelbrot<sup>29</sup>, Falconer<sup>30,31</sup>, Barnsley<sup>3</sup> and Chin et al.<sup>32</sup> for instance. The approach undertakes an "inverse problem", that is, the determination of the IFS underlying a particular fractal, in this case the electroencephalographic signal. The procedure proposed here uses a special family of mappings whose attractor is the graph of a function fitting the data.

In the case of an affine IFS, no additional numerical procedure is needed because there are explicit and exact formulae providing the dimension in terms of the IFS coefficients (Barnsley<sup>3</sup>, Falconer<sup>33</sup>). These numbers provide numerical characterizations of the geometric complexity of the signals. The computational cost of the process is almost zero for short recordings. The method does not involve problems of convergence or the choice of an excessively large number of algorithmic parameters. It can be used to quantify series of any length. The only hypotheses are the continuity and the characteristic selfsimilarity of the fractal interpolation functions.

On the other hand, our results suggest that the fractal dimension is as useful as the embedding dimensions in order to record the variations of the EEG complexity inherent to the bioelectric changes produced by metabolic changes in the cerebral

areas implied in a specific mental process and/or pathology<sup>10,11</sup>.

## Acknowledgments

This paper is part of a research project financed by the Consejo Superior de Investigación y Desarrollo de la Diputación General de Aragón (P072/2001) and by the University of Zaragoza (Spain) (226-55).

## References

1. M.F. Barnsley, *Constr. Approx.* **2**(4), 303 (1986).
2. M.A. Navascués and M.V. Sebastián, *Fractals* **11**(1), 1 (2003).
3. M.F. Barnsley, *Fractals Everywhere*, (Academic Press, Inc, 1988).
4. E.W. Cheney, *Approximation Theory*, (AMS Chelsea Publ. 1966).
5. W.C. Strahle, *AIAA J.* **29**(3), 409 (1991).
6. A. Babloyantz, J.M. Salazar and C. Nicolis, *Phys. Lett.* **111A** (3), 152 (1985).
7. K.J. Stam, D.L. Tavy, B. Jelles, H.A. Achtereekte, J.P. Slaets and R.W. Kenunen, *Brain Topography* **7**(2), 141 (1994).
8. X. Nan and X. Jinghua, *Bull. Math. Biol.* **50**, 559 (1988).
9. A.S. Besicovitch and H.D. Ursell, *J. of London Math. Soc.* **95**, 263 (1937).
10. J.R. Valdizán, M.A. Navascués and M.V. Sebastián, *Rev. Neurol.* **25**(148), 1882 (1997).
11. J.R. Valdizán, M.A. Navascués and M.V. Sebastián, *Rev. Neurol.* **32**, 127 (2001).
12. K.E. Graf and T. Elbert in *Brain Dynamics*, ed. E. Basar, (Springer-Verlag, 1989).
13. W.S. Pritchard and D.W. Duke, *Psychophysiology* **27**, 56 (1990).
14. R.A.M. Gregson, L.A. Britton, E.A. Campbell and R.G. Gates, *Biol. Psychol.* **31**, 173 (1990).
15. T.S. Parker and L.O. Chua, *Practical Numerical Algorithms for Chaotic Systems*, (Springer Verlag, 1989).
16. A. Renyi, *Probability Theory*, (North-Holland, 1970).
17. F. Takens in: *Dynamical Systems and Turbulence*, eds. A. Rand and L.S. Young, (Springer Verlag, 1981).
18. P. Grassberger and I. Procaccia, *Phys. Rev. Lett.* **50**, 346 (1983).
19. G. Mayer-Kress (ed.), *Dimensions and Entropies in Chaotic Systems* (Springer Verlag, 1986).
20. L.A. Smith, *Phys. Lett. A* **133-6**, 283 (1988).
21. E. Basar, C. Basar-Eroglu, J. Röschke and J. Shult in: *Models of Brain Function*, ed. R.M.J. Cotterill, (Cambridge Univ. Press, 1989).
22. F.H. Lopes da Silva, W. Kamphuis, J.M.A.M. van Neerven, J.P.M. Pijn in: *Machinery of Mind*, ed. E.R. John, (Birkhäuser, 1990).
23. M. Ding, C. Grebogi, E. Ott, T. Sauer and J. A. Yorke. *Phys. Rev. Lett.* **70**(25), 3872 (1993).
24. S. Ellner, *Phys. Lett. A* **133**(3), 128 (1988).

25. J. Holzfuss and G. Mayer-Kress, in *Dimensions and Entropies in Chaotic Systems*, ed. G. Mayer-Kress, (Springer Verlag, 1986).
26. D.S. Broomhead and G.P. King, *Physica* **20D**, 217 (1986).
27. J.H. Friedman, W. Stuetzle and A. Schroeder, *J. of the Am. Stat. Assoc.* **79**(387), 599 (1984).
28. A.I. Mees, P.E. Rapp and L.S. Jennings, *Phys. Rev. A* **36**(1), 340 (1987).
29. B.B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, 1977).
30. K.J. Falconer, *Math. Proc. Camb. Phil. Soc.* **103**, 339 (1988).
31. K.J. Falconer, *Fractal Geometry, Mathematical Foundations and Applications*, (John Wiley & Sons, 1990).
32. W. Chin, B. Hunt and J.A. Yorke, *Trans. Am. Math. Soc.* **349**, 1783 (1997).
33. K.J. Falconer, *J. Stat. Phys.* **47**, 123 (1987).

# STOCHASTIC AND REGULAR COMPONENTS IN FORCING OF SOLAR LARGE-SCALE STRUCTURES

E. TIKHOMOLOV

*TRIUMF, Canada's National Laboratory for Particle and Nuclear Physics, 4004  
Wesbrook Mall, Vancouver, BC, V6T 2A3, Canada  
E-mail: etikhomolov@solar.stanford.edu*

Large-scale organization of the solar magnetic fields is considered to be the result of forcing of large-scale vortices at the bottom of the convection zone. One of the sources of such vortices is the penetrative convection that seeds inverse cascade from small to large scales. The other source is deformational long-wave instability that directly amplifies the large-scale disturbances and forces differential rotation. The combined action of these two processes is numerically simulated. It is shown how the large-scale vortical pattern is reflected in the distribution of the magnetic elements. Large-scale vortices have much longer lifetime than small magnetic elements. Such effects can be used for interpretation of such phenomena as solar complexes of activity and fractal properties of the large-scale magnetic fields.

## 1 Introduction

The observations of the sun revealed that solar magnetic fields demonstrate an enclosing hierarchy in their spatial organization: the large magnetic elements consist of smaller ones. The largest size is on the order of solar radius. At this top level small-scale short-lived (with lifetime on the order of one day) magnetic elements are organized into large-scale long-lived (with lifetime up to several solar rotations) active regions and complexes of activity. Another demonstration of large-scale organization is quasi-periodic rearrangement of the magnetic field pattern known as the 11-year solar cycle <sup>1</sup>. At present, the lowest limit for the scale of spatial organization is unknown. The investigation of possible reasons for such organization is very important for forecasting the solar activity.

The existence of spatial hierarchy in solar magnetic elements attracted attention to their fractal properties and reasons of their organization. Fractal analysis gives value to fractal dimension of magnetic elements in the range:  $1 < d < 1.8$  <sup>2,3</sup>. One of the interpretations of large-scale organization is self-organization of the magnetic elements that interact with each other <sup>4,5</sup>. We argued <sup>6</sup> that this approach can be used only for the explanation of relatively small formations, such as groups of sunspots, but not of complexes of activity or active longitudes.

From our point of view, more promising is the interpretation of organization and fractal properties of magnetic fields as the result of excitation of large-scale vortical flows at the base of the solar convection zone and generation of the magnetic field by them. In our earlier publications, <sup>7</sup> we investigated how large-scale vortices generate small-scale magnetic elements because of twisting of the magnetic field lines and subsequent reconnection. The process of reconnection leads to formation of small-scale magnetic elements that can reflect the velocity field of flows inside a separate vortex. At the same time reconnection can lead to stochastization in distribution of magnetic elements.

The source of large-scale vortices at the bottom of the convection zone is consid-

ered to be the penetrative convection that works as a stochastic source of the initial disturbances. These initial disturbances merge into large-scale vortices, demonstrating inverse cascade. This process was numerically simulated in application to the problem of Great Red Spot of Jupiter <sup>8</sup> and recently in application to the sun <sup>9,10</sup>. Stochastic forcing of large-scale vortices leads to stochastic distribution of large-scale patterns of magnetic elements. On the scale of solar radius the enhanced concentration of magnetic elements inside vortices appears as the large-scale organization of the magnetic field with a characteristic lifetime of a large-scale vortex.

Recently, we developed a hydrodynamical model of 11-year solar variations in which oscillations appear as the result of interaction between shear and deformational long-wave instability (DLWI) <sup>11</sup>. At this top level of organization the magnetic field generated due to 11-year hydrodynamic oscillations has 22-year period and is the seed field amplified by the large-scale vortices.

The formation of large-scale vortical patterns should be considered along with the effects leading to 11-year variations because of the possible interaction between these two processes. Stochastic forcing of vortical flows and related inverse cascade can significantly change the conditions for excitation of 11-year oscillations. On the other hand DLWI can directly force the large-scale vortical flows and change the conditions for formation of large-scale vortices. The goal of this paper is the investigation of interaction between stochastic and regular components of large-scale hydrodynamic flows and its reflection on the top level of organization of the solar magnetic field.

## 2 Model

The layer of forcing of large-scale vortices is supposed to be located below the solar convection zone in the penetrative convection region. This layer is considered in shallow-water and beta-plane approximations. The lower and upper boundaries of the layer are presumed to be, respectively, deformed and nondeformed free surfaces. The deformation of upper surface of the layer is associated with a perturbation of the isopicnic surfaces near the bottom of the convection zone that appears when the flows are excited. At the boundaries corresponding to the poles, the velocity is specified to be zero. The equations are written in a Cartesian coordinate system, which is rotating with the velocity of plasma at latitude  $30^\circ$ . Established results of helioseismology show that the upper part of the radiative zone rotates with this velocity at all latitudes. For all magnetic field components on the lower boundary of the layer, the condition of perfect conductivity is specified. On the side boundaries (corresponding to polar regions) conditions are specified in such a way that the magnetic field is parallel to the polar axis. On the upper surface of the layer, for the horizontal components of the magnetic field, the condition of perfect conductivity is assumed, and for the vertical component an "open" boundary condition is used <sup>12</sup> (i.e. it is supposed that this component freely emerges through the upper surface of the layer).

The following units are used: for the horizontal coordinates  $x$  and  $y$  – the size of the order of one third of the solar radius  $R \approx 2 \times 10^{10}$  cm; for the vertical coordinate  $z$  – the thickness of the layer of Rossby vortex excitation  $h_0$ ; we assume

$h_0 \approx 10^9$  cm; and for time  $-t_0 = 1$  year  $\approx 13$  solar rotations. In these units, the dimensionless equation for the stream function  $\psi$  has the form:

$$\begin{aligned} \psi_t/Q - \Delta\psi_t - YJ(\psi, \Delta\psi) - B\psi_x + D_v\Delta^2\psi + \gamma\Delta\psi &= 0, \\ \Delta &= \partial^2/\partial x^2 + \partial^2/\partial y^2, \quad J(f, g) = f_x g_y - f_y g_x, \end{aligned} \quad (1)$$

where  $Q = (r_R/R)^2$  is the Froude number,  $r_R = (gh_0)^{1/2}/2\Omega$  is Rossby-Obukhov deformation radius,  $g$  is gravity,  $\Omega$  is the angular rotation velocity,  $Y = 2\Omega t_0 Q$ ,  $B$  is the parameter characterizing the beta-effect, the dependence of the angular velocity on latitude. For the chosen units and parameters, we have  $Q \approx 1$ , and  $Y \approx 150$ . The parameter  $B$  is taken to be  $B = 10$ .

The diffusion coefficient has the form  $D_v = D_0 + W^n \exp(-((y - y_{n60})/\delta)^2) + W^s \exp(-((y - y_{s60})/\delta)^2)$ .  $W^n, W^s$  are the amplitudes of small-scale vortical flows which are assumed to be excited by virtue of shear instability in each hemisphere in the regions centered on latitudes  $y_{n60} = 60^\circ N$ ,  $y_{s60} = 60^\circ S$ . The width of these regions is chosen as  $\delta = 70^\circ$  in equatorward direction and  $\delta = 15^\circ$  in poleward direction.  $D_0 = \nu t_0/R^2$ , where  $\nu$  is the coefficient of effective turbulent kinematic viscosity which we take in the penetrative convection region to be the order of  $10^{12} \text{ cm}^2 \text{ s}^{-1}$ ; hence  $D_0 = 0.1$ . Introducing the dependence of the coefficient of positive diffusivity  $D_v$  on the amplitude of vortical flows in the suggested simple form seems rather natural, because the effective turbulent viscosity increases proportionally to the amplitude of the vortical flows. Deformation of the interfacial surface dividing convectively stable and convectively unstable parts of the layer in the northern and southern hemispheres is derived, respectively, according to the formulas  $h = \psi$  and  $h = -\psi$ .

The coefficient of negative diffusion in each hemisphere is taken in the form  $\gamma = \gamma_0(\exp(-((y - y_{n60})/\eta)^2) + \exp(-((y - y_{s60})/\eta)^2))$ .  $\gamma_0$  is taken to be 0.5. The width is chosen as  $\eta = 40^\circ$  in equatorward direction and  $\eta = 15^\circ$  in poleward direction.

In our previous publications, we studied a simpler case, when the parameters  $D_v$  and  $\gamma$  were constants<sup>13</sup>. Under this condition, the amplitude increases for solutions having the wave vector less than  $k_{cr} = (\gamma/D_v)^{1/2}$ . In the case considered here, there is no simple analytical formula for a critical wave vector because of the appearance of a dependence of the positive diffusion coefficient on amplitude of vortical flows and on the latitude, and also because of the appearance of a dependence of negative diffusion coefficient on the latitude. It is apparent, however, that an increase in amplitude of vortical flows leading to an increase of the positive diffusion coefficient finally brings about disruption of forcing of zonal flow.

A mathematical formulation of the statement that a shear instability is realized in high latitude regions is the Ginzburg-Landau equation derived in weakly-nonlinear theory and describing the dynamics of vortical flows in these regions<sup>14</sup>. In the dimensionless form for north and south hemispheres, it reads as follows:

$$\begin{aligned} W_t^n &= \lambda^n W^n - (W^n)^3 \\ W_t^s &= \lambda^s W^s - (W^s)^3. \end{aligned} \quad (2)$$

$W^n, W^s$  are considered to be real, nonnegative functions. The role of the cubic term is the limitation on the amplitude of vortical flows. The first term on the right-

hand side of Eq. (2) describes an increase in the disturbance amplitude provided by a shear instability.

A well-known condition for shear instability development in an ideal non-viscous fluid is fulfilment of the Rayleigh instability criteria: A change of sign of the second derivative of velocity<sup>14</sup>. In a non-ideal fluid, for the excitation of vortical flows it is also necessary that the Reynolds number (defined by the characteristic flow velocity under condition of specified geometry and viscosity) has a rather large value. In addition, it is also necessary to take into consideration the hysteresis phenomenon: Reynolds number is larger in the case when instability begins than when it stops. Thus, for modelling the critical character of the excitation of vortical flows it is taken:

$$\begin{aligned}\lambda^n &= 1, \text{ if } u^n < u_{\min}, \\ \lambda^n &= -1, \text{ if } u^n > u_{\max}, \\ \lambda^s &= 1, \text{ if } u^s < u_{\min}, \\ \lambda^s &= -1, \text{ if } u^s > u_{\max},\end{aligned}\quad (3)$$

where we consider to be critical values  $u_{\min} = -140 \text{ m s}^{-1}$  and  $u_{\max} = -115 \text{ m s}^{-1}$ . Hence, Eq. (2) describe an increase in amplitude of vortical flows as zonal velocity attains a critical value  $u_{\min}$ , and decrease in amplitude as zonal velocity becomes greater than  $u_{\max}$ .

The equations governing the magnetic field evolution are used in a heliomagnetostrophic approximation<sup>15</sup>. The horizontal components of the magnetic field are represented as the sums of toroidal and poloidal components:  $B^x = -T_y + P_x$ ,  $B^y = T_x + P_y$ . The vertical component of the magnetic field at the upper surface of the layer is inferred by  $B^z = -\Delta P$ . The unit of the horizontal components of the magnetic field  $B^x$  and  $B^y$  is the characteristic value of the magnetic field strength  $M$ , while the unit of the vertical component  $B^z$  is  $\kappa M$ , where  $\kappa = h_0/R$ .

The equations for the toroidal  $T$  and the poloidal  $P$  functions have the form<sup>7,16</sup>:

$$\Delta T_t = -\Delta(uT_x + vT_y) - \Delta(uP_y - vP_x) - (v\Delta P)_x + (u\Delta P)_y + D_m\Delta^2 T, \quad (4)$$

$$\Delta P_t = -u\Delta P_x - v\Delta P_y + w_x T_y - w_y T_x + \alpha(T_{xx} + T_{yy}) + D_m\Delta^2 P, \quad (5)$$

where  $D_m = \mu t_0/R^2$ , and  $\mu \approx \nu$  is the effective coefficient of turbulent magnetic diffusion. So  $D_m \approx D_v \approx 0.1$ . In this paper we neglect the back influence of the magnetic field upon hydrodynamic flows using the approach employed in kinematic dynamo theory. The velocity components that are substituted into the equations for the magnetic field are determined, respectively, by:  $u = -Y\psi_y$ ,  $v = Y\psi_x$ ,  $w = \psi_t$ .

We parametrize vortical flows excited by virtue of shear instability. Under the condition of nonzero mean helicity, these vortical flows give rise to a well-known  $\alpha$ -effect<sup>17</sup> which is described by adding to the right-hand side of equation (5) the term of the form  $\alpha(T_{xx} + T_{yy})$ . One can expect that  $\alpha$ -effect arises only under the condition of rather large amplitude of vortical flow and is suppressed when magnetic field strength attains large values. Considering these effects and the limited action

of  $\alpha$ -effect in space,  $\alpha$  is chosen in the form:

$$\alpha = \alpha_0((W^n - W_{cr})(1 + \tanh(\sigma(W^n - W_{cr}))) \exp(-((y - y_{n60})/\delta)^2) + (W^s - W_{cr})(1 + \tanh(\sigma(W^s - W_{cr}))) \exp(-((y - y_{s60})/\delta)^2))/2(1 + |\mathbf{B}|^2) \quad (6)$$

with  $\alpha_0 = 0.4$ ,  $\delta = 15^\circ$ , and  $W_{cr} = 0.6$ .

Parameter  $\sigma$  determines the rate of "turning on" the  $\alpha$ -effect and is specified to be 10. Under the condition considered, an amplification of the magnetic field strength takes place. To limit the amplitude, the following well-known method is used: the magnetic field strength squared is introduced as the denominator into the formula for  $\alpha$ .

To include geometrical effects in this paper we introduce a limitation on the model: in the latitudes higher than  $75^\circ$ , near  $45^\circ$ , and in the equatorial region  $5^\circ S - 5^\circ N$  flows are assumed to be axisymmetric. To simulate this situation, we use the technique of reducing down to zero the coefficients of all components in the equations that are dependent on longitude.

Penetrative convection is simulated by adding a small single Gaussian vortex to the flow at time intervals of one twentieth of the Carrington rotation period. The location, size and deformation of the upper boundary (associated with velocity) are specified randomly. Values vary in the range of  $0^\circ - 360^\circ$  for location in longitude,  $75^\circ S - 75^\circ N$  for location in latitude,  $1^\circ - 5^\circ$  for size, and  $0 - 5 \times 10^7$  cm for deformation.

### 3 Results

Partial differential equations described in previous section are solved numerically using semi-spectral method. 11-year hydrodynamic oscillations in high latitudes give rise to 22-year magnetic oscillations. At the same time penetrative convection disturbances merge into vortices of different scales demonstrating inverse cascade.

#### 3.1 Quasi-regular component

Figure 1 shows dynamics of the quasi-regular component. At the top of Fig.1, the oscillations of zonal flows are shown. Frame of reference rotates with the velocity of plasma at  $30^\circ$ , thus zonal velocity has positive and negative values. In high and mid latitudes zonal flow has the direction from west to east. At the equator the flow has the maximum amplitude and shear has cyclonic character in accordance with observations. 11-year oscillations of zonal flow appear because of competition between forcing by DLWI and suppressing due to increase of effective turbulent diffusion.

The oscillations of the magnetic field component are shown in the middle and at the bottom of Fig.1. One can see time lag between maximums of toroidal and vertical magnetic field components. In our model, the period of magnetic oscillations is determined solely by hydrodynamic factors, and 22-year oscillations are excited as a result of the impulsive development of shear instability every 11-year and the corresponding impulsive excitation of  $\alpha$ -effect. Our numerical simulations



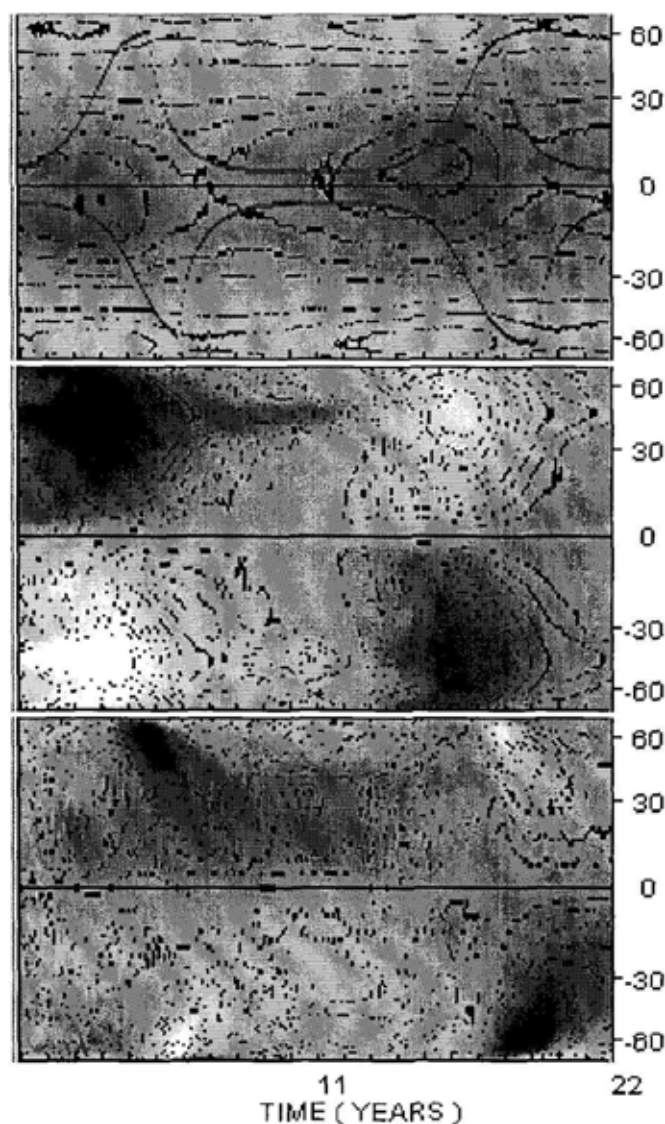


Figure 1. 11-year hydrodynamic and 22-year magnetic oscillations. Top: 11-year oscillations of zonal flow. Solid lines show oscillations of diffusion coefficient  $W$  at latitudes  $60^\circ$  for south and north hemispheres. Middle: oscillations of the toroidal component. Bottom: oscillations of the vertical component of the magnetic field. Minimum negative and maximum positive values correspond, respectively, to white and black color.

show that localization of  $\alpha$ -effect in high latitudes does not preclude the formation of a strong toroidal magnetic field in the low latitudes.

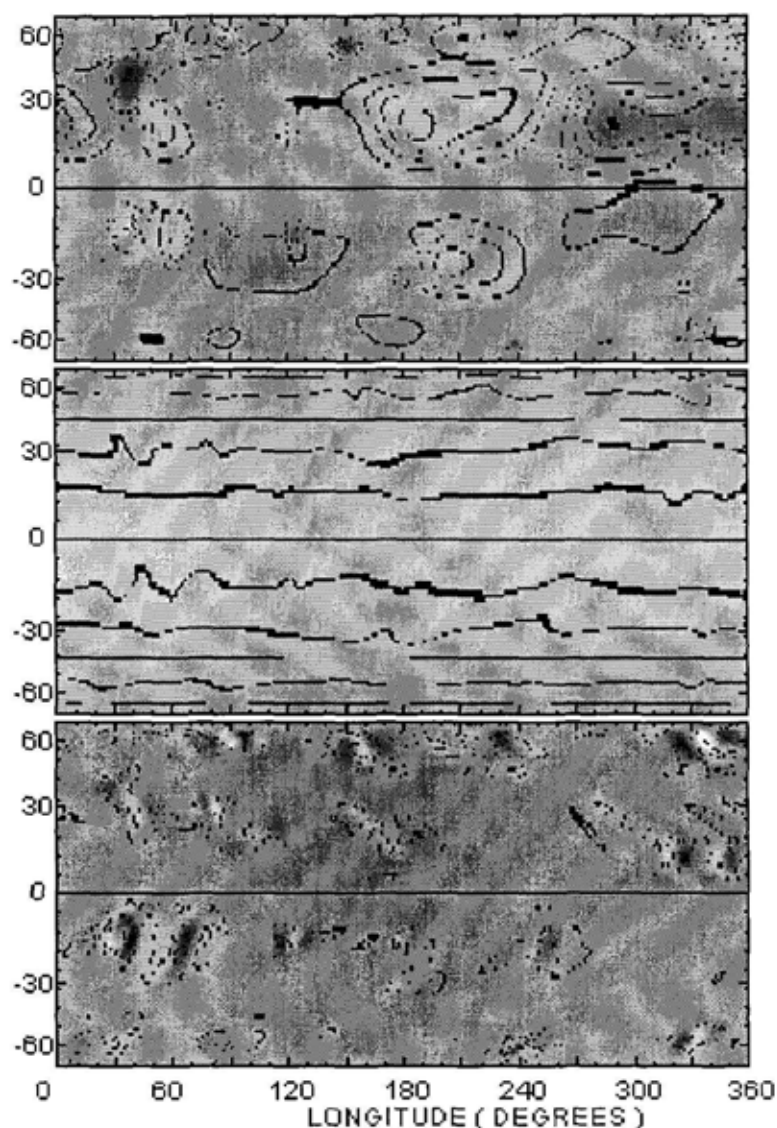


Figure 2. Distribution of vortices and the magnetic field at the maximum of solar cycle. Top: distribution of stream function  $\psi$ . Middle: distribution of toroidal component  $T$  function. Bottom: distribution of vertical component  $B^2$ . Minimum negative and maximum positive values correspond, respectively, to white and black color.

### 3.2 Stochastic component

One of the main processes for stochastic component is the merging of the initially excited vortices. The stationary distribution of vortices after several hundred years is shown at the top of Fig. 2 ( differential rotation is subtracted in this figure).

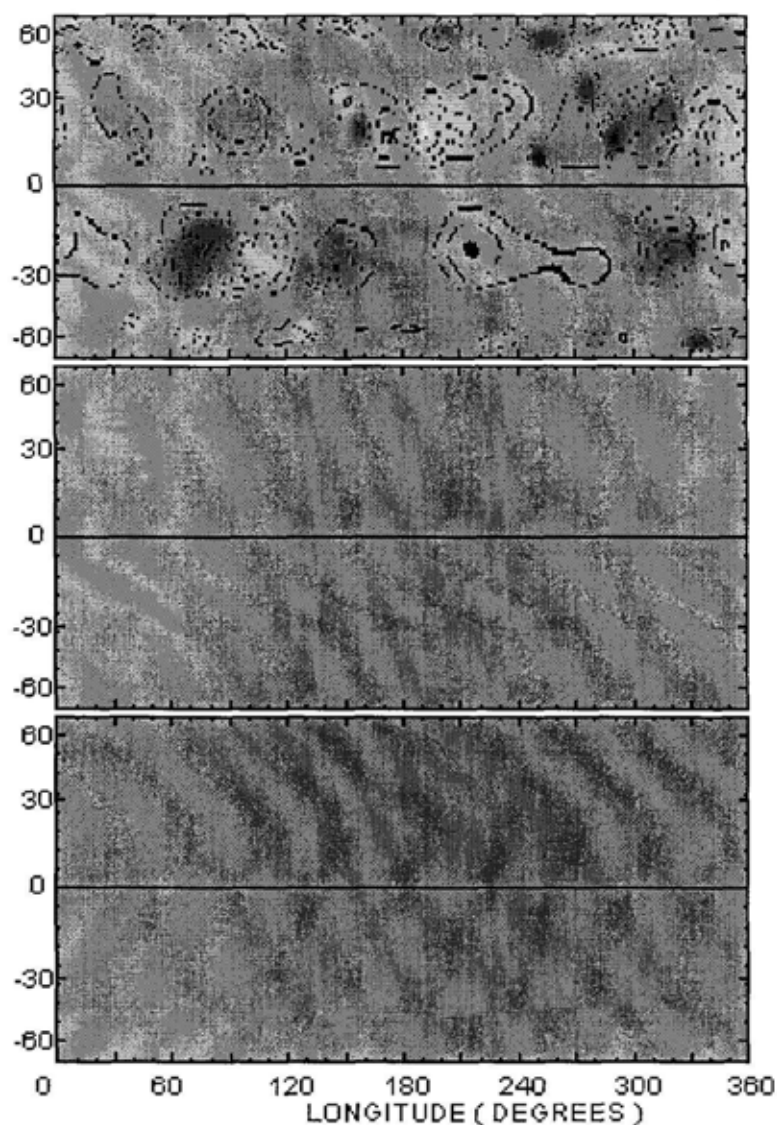


Figure 3. Distribution of vortices and the magnetic field at the minimum of solar cycle. Top: distribution of stream function  $\psi$ . Middle: distribution of toroidal component  $T$  function. Bottom: distribution of vertical component  $B^2$ . Minimum negative and maximum positive values correspond, respectively, to white and black color.

At high latitudes vortices are forced both by penetrative convection and DWLI. Magnetic field generated by quasi-regular component is a seed field for vortices. Vortices twist magnetic field lines and magnetic field is amplified inside vortices. This process is seen in the middle of Fig.2. Differential rotation stretches the magnetic field lines in east-west direction that leads to appearance of some angle

between axis of magnetic elements and the equator. This effect is seen at the bottom of Fig.2 as the arc-like (or banana-like) shape distribution of the magnetic field.

Amplification of the magnetic field can lead to the initialization of transfer to the surface of the magnetic field due to magnetic buoyancy. Figure 2 shows the flows and magnetic field at the maximum of solar cycle when toroidal quasi-regular magnetic field component has maximum amplitude.

Vortices have various spatial scales. Large-scale vortices have a lifetime much longer than small magnetic elements. This effect can be associated with such phenomena as complexes of activity considering that sources of complexes of activity are the large-scale vortices.

Figure 3 shows the flows and the magnetic field at the minimum of solar cycle when toroidal quasi-regular magnetic field component has minimum amplitude. Strong magnetic fields entirely disappear. However, vortices continue to exist. In such situation the "new" strong magnetic field of the next cycle can be generated by the same large-scale vortices that generated the "old" strong magnetic field. Thus in the frame of our approach we can interpret such long-term phenomena as solar active longitudes.

Stochastic component does not preclude the excitation of 11-year hydrodynamic and 22-year magnetic oscillations. Comparison with the results of our previous paper <sup>11</sup> shows that it adds new features to the dynamics of zonal flows and axisymmetric component of the magnetic field. One can see in Fig.1 the effect of slight asynchronization of oscillations in south and north hemispheres. Such phenomenon is observed on the sun. Another effect is short variations in distributions of vertical magnetic field component seen at the bottom of Fig 1. All these effects are related to forcing of vortices by penetrative convection.

#### 4 Discussion

An evolution of the large-scale solar magnetic fields shows an existence of quasi-regular and stochastic components. In standard solar dynamo models quasi-regular component oscillating with 11-year (or 22 year for the magnetic field) period is considered to be the result of action of turbulence with nonzero mean helicity. Period of oscillations is established due to the structure of the equations for the magnetic field. The large-scale magnetic structures are interpreted as the result of excitation of non-axisymmetric dynamo component. However, such approach is not able to interpret a number of phenomena observed for the complexes of activity and active longitudes. Numerical simulations of global convection showed that lifetime of separate convective cell is on the order of period of one solar rotation (see, for example, results of Gilman and Miller <sup>18</sup>), hence the convection cells cannot be considered as the sources of global magnetic structures and complexes of activity.

The presented approach moves the problems of quasi-regular and stochastic components into hydrodynamic domain: large-scale magnetic field dynamics is a reflection of the hydrodynamical processes at the base of the convection zone. The source of stochastic component is penetrative convection that forces large-scale vortical flows. The regular component appears as a result of hydrodynamic 11-year

oscillations. As the result the large-scale magnetic field generated due to the both processes has stochastic and regular components. Within such approach it is possible to explain such phenomena as 11-year variations, complexes of activity and active longitudes on a common "hydrodynamic" basis.

We also suggest a way for the interpretation of fractal properties of the solar magnetic field. Stochastic forcing of the large-scale vortices, their merging, and interaction with a flow oscillating with 11-year period create a background on which a generation of magnetic field occurs. Twisting of the magnetic field line by vortices and reconnection of magnetic field lines on the small and large scales lead to appearance of hierarchy in their space organization. Our future plan is to do numerical simulations with higher resolution that will cover at least three levels of magnetic field organization. A comparison of fractal dimension of the generated magnetic elements with obtained from observations can be a good test for our theory of organization of solar magnetic fields.

## References

1. E.R. Priest, *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982).
2. Z. Mouradian, I. Soru-Escout, *Astron. Astrophys.* **251**, 649 (1991).
3. N. Meunier, *Astrophys. J.* **515**, 801 (1999).
4. E.I. Mogilevsky *et al* in *Contr. Astron. Obs. Skalnaté Pleso, Proc. 12th Region. Consult. on Solar Phys.*, Vol. 15, p. 189, ed. A. Antalova, (Astron. Ustav Slovenskej Akad. Vied, Tatranska Lomnica, 1986).
5. L. Vlahos, T. Fragos, H. Isliker, M. Georgoulis *Astrophys. J.* **575**, L87 (2002).
6. E. Tikhomolov, *Astron. Nachr.* **319**, 245 (1998).
7. E. Tikhomolov, *Solar Phys.* **156**, 205 (1995).
8. G.P. Williams, R.J. Wilson, *J. Atmos. Sci.* **45**, 207 (1988).
9. E. Tikhomolov, V. Mordvinov, *Astron. Nachr.* **322**, 189 (2001).
10. E. Tikhomolov in *Current Theoretical Models and High Resolution Solar Observations: Preparing for ATST. ASP Conference Series*, eds. A.A. Pevtsov, H. Uitenbroek, Vol. 286, pp. 113-120 (The Astronomical Society of the Pacific, San Francisco, 2003).
11. E. Tikhomolov, *Solar Phys.* **199**, 165 (2001).
12. H. Yoshimura, *Astrophys. J.* **247**, 1102 (1981).
13. E. Tikhomolov, *Physics Fluids* **8**, 3329 (1996).
14. L.D. Landau, E.M. Lifshitz, *Hydrodynamics* (Nauka, Moscow, 1988).
15. P.A. Gilman, *Solar Phys.* **8**, 316 (1969).
16. E. Tikhomolov, V. Mordvinov, *Astrophys. J.* **472**, 389 (1996).
17. E. Parker, *Cosmical Magnetic Fields* (Clarendon Press, Oxford, 1979).
18. P.A. Gilman, J. Miller, *Astrophys. J. Suppl.* **61**, 585 (1986).

# FAST, EFFICIENT ON-LINE SIMULATION OF SELF-SIMILAR PROCESSES

OWEN DAFYDD JONES

*School of Mathematics, University of Southampton, Highfield, Southampton SO17 1BJ,  
U.K.*

*E-mail: O.D.Jones@maths.soton.ac.uk*

We describe a class of self-similar processes that can be used to fit self-similar data, and give a fast, efficient on-line algorithm for simulating them.

## 1 Introduction

Self-similar processes are of interest as models for internet packet arrival data, high-frequency financial data, ECG and EEG traces, and various hydrological and meteorological time series. Simulation of self-similar processes has proven problematic, because they exhibit a slowly decaying correlation structure (long-range dependence), which means that the individual elements of any sequence of observations  $X(1), \dots, X(n)$  are strongly correlated. In practice to date this generally means that it is not possible to simulate  $X(n)$  without simultaneously simulating  $X(1), \dots, X(n-1)$ , and this necessarily results in an algorithm that requires  $O(n)$  storage to generate  $n$  steps of  $X$ . More importantly, this means that if you have already generated  $n$  steps, then it is not possible to generate step  $n+1$  directly, instead it is necessary to generate all  $n+1$  steps from scratch.<sup>1,2,3</sup>

One model which avoids these problems is the  $M/G/\infty$  queue.<sup>4</sup> Unfortunately this model is not flexible enough to be useful in practice. Here we present a new class of self-similar models called EBP processes (for Embedded Branching Process), which are flexible, readily fitted to data, and easily simulated. Features of the simulation algorithm are

- (i) Scaleable:  $O(n \log n)$  time to generate  $n$  steps.
- (ii) Efficient:  $O(\log n)$  storage required to generate  $n$  steps.
- (iii) On line: can generate a new step on demand.

The class of processes is described in Section 2, and the simulation algorithm given in Section 3. A `matlab` implementation of the algorithm can be found on the authors web page at [www.maths.soton.ac.uk/staff/ODJones/](http://www.maths.soton.ac.uk/staff/ODJones/). Some examples are given in Section 3.

## 2 EBP processes

Suppose that  $X : \mathbb{R}_+ \rightarrow \mathbb{R}$  is a continuous process.  $X$  is said to be self-similar if for some  $H$  and all  $a > 0$

$$X(t) = a^{-H} X(at) \text{ in distribution.}$$

$H$  is called the *Hurst index*.

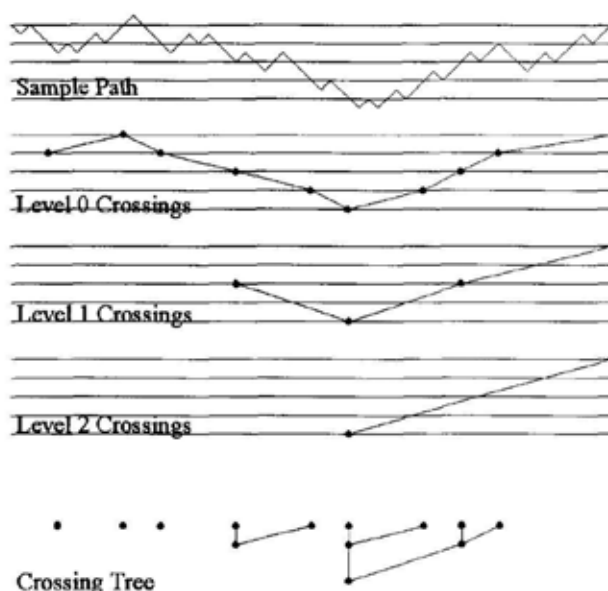


Figure 1. An example of a crossing tree.

We construct the diadic crossing tree for  $X$ . W.l.o.g. suppose  $X(0) = 0$ . For any  $n \in \mathbb{Z}$ , let  $T_0^n = 0$  and  $T_{k+1}^n = \inf\{t > T_k^n : X(t) \in 2^n\mathbb{Z}, X(t) \neq X(T_k^n)\}$  be the hitting times of the size  $2^n$  crossings of the process. The path from  $X(0)$  to  $X(T_1^n)$  is generally not a true crossing, so we discard it. That is, the  $k$ -th size  $2^n$  crossing is from  $X(T_k^n)$  to  $X(T_{k+1}^n)$ .

There is a natural tree structure to the crossings, as each crossing of size  $2^n$  can be decomposed into a sequence of crossings of size  $2^{n-1}$ . The nodes of the crossing tree are crossings, and the offspring of any given crossing are the corresponding set of subcrossings at the level below. Let  $Z_k^n$  be the number of subcrossings of size  $2^{n-1}$  that make up the  $k$ -th crossing of size  $2^n$ . A crossing tree is illustrated in Fig. 1. Note that the crossing tree is well defined for any continuous process, not just self similar processes.

If  $X$  is self-similar with stationary increments, then it can be shown that the  $Z_k^n$  form a stationary sequence. Conversely, we will call any continuous process  $X$  an *Embedded-Branching-Process* (EBP) process if the  $Z_k^n$  are independent and identically distributed. In this case the tree descending from any fixed crossing is a realisation of a Galton-Watson branching process.

From now on let  $X(t)$  be an EBP process, and let  $p(x) = P(Z_k^n = x)$  be the *offspring distribution*. Clearly  $p$  satisfies  $\{x : p(x) > 0\} \subset \{2, 4, 6, \dots\}$ . If in addition we have

$$p(2) < 1 \text{ and } \sum_{x=1}^{\infty} x \log(x) p(x) < \infty,$$

then we say that  $p$  is *regular*.

**Theorem 1** *For any regular discrete distribution  $p$  there exists a continuous EBP process  $X$  for which  $p$  is the offspring distribution. Let  $\mu = \sum_{x=1}^{\infty} x p(x)$  and  $H = \log 2 / \log \mu$ , then for all  $a = \mu^n$ ,  $n \in \mathbb{Z}$*

$$X(t) = a^{-H} X(at) \text{ in distribution.} \quad (1)$$

The proof is deferred to Section 4.

## 2.1 Fitting to data

Fitting an EBP process requires an estimate of the subcrossing distribution  $p$ . This is readily achieved by computing the crossing tree, and then using the observed subcrossing numbers  $Z_k^n$  to estimate  $p$ . One can easily test the assumption that the  $Z_k^n$  are i.i.d., and in practice this is seen to be a reasonable assumption to make for self-similar data. Thus EBP processes form an extremely flexible class of models for self-similar processes.

An application of the crossing tree to the estimation of the Hurst index  $H$  of a self-similar process is given by Jones & Shen.<sup>5</sup>

## 2.2 Markov representation

Let  $X^m$  be the random walk on  $2^m \mathbb{Z}$  obtained by observing only  $2^m$  crossings of  $X$ . That is  $X^m(k) = X(T_k^m)$  for  $k = 1, 2, \dots$ . In this section we give an infinite dimensional representation of  $X^m$  which is Markov.

Let  $C_k^n$  be the  $k$ -th crossing of size  $2^n$ . By a crossing we mean a section of the sample path, plus some extra information such as the time and place the crossing starts. We adopt the convention that a crossing includes its initial point but not its final point. For  $n \geq m$  let  $\kappa(m, n, k)$  be such that  $X^m(k) \in C_{\kappa(m, n, k)}^n$ , for  $k = 1, 2, \dots$ . We have that  $C_k^n$  has  $Z_k^n$  subcrossings, and define  $S_k^n$  be such that  $C_k^n$  is subcrossing number  $S_k^n$  of  $C_{\kappa(n, n+1, k)}^{n+1}$ . Clearly  $1 \leq S_k^n \leq Z_{\kappa(n, n+1, k)}^{n+1}$ .

A crossing has one of 6 types depending upon its direction (up or down) and where it starts from. Suppose that we have a crossing of size  $2^n$ , and that the parent crossing starts from  $k2^{n+1}$ . The 6 types are then  $0^+, 0^-, 1^+, 1^-, -1^+, -1^-$ , where a type  $i^+$  crossing is from  $k2^{n+1} + i2^n$  to  $k2^{n+1} + (i+1)2^n$ , and a type  $i^-$  crossing is from  $k2^{n+1} + i2^n$  to  $k2^{n+1} + (i-1)2^n$ . Let  $\alpha_k^n$  be the type of crossing  $C_k^n$ .

We define the *crossing state* of  $X^m$  at time  $k$  to be  $\mathcal{X}^m(k) = \{\mathcal{X}^{m,n}(k)\}_{n \geq m}$  where

$$\mathcal{X}^{m,n}(k) = (\kappa(m, n, k), S_{\kappa(m, n, k)}^n, Z_{\kappa(m, n+1, k)}^{n+1}, \alpha_{\kappa(m, n, k)}^n).$$

We will occasionally write  $S^{m,n}(k)$ ,  $Z^{m,n+1}(k)$  and  $\alpha^{m,n}(k)$  for  $S_{\kappa(m, n, k)}^n$ ,  $Z_{\kappa(m, n+1, k)}^{n+1}$  and  $\alpha_{\kappa(m, n, k)}^n$ . If  $S^{m,n}(k) = Z^{m,n+1}(k)$  then  $X^m(k)$  is at the end of a level  $n+1$  crossing.

**Theorem 2**  $\mathcal{X}^m$  is a Markov chain.



**Proof** We describe how to generate  $\mathcal{X}^m(k+1)$  from  $\mathcal{X}^m(k)$  using the recursive procedure **Increment** acting on the  $\mathcal{X}^{m,n}(k)$ . **Increment** is applied to  $\mathcal{X}^{m,n}(k)$  when step  $k+1$  of  $\mathcal{X}^m$  takes it into a new crossing at level  $n$ .

```

Procedure Increment  $\mathcal{X}^{m,n}(k)$ 
  (Assume that  $\mathcal{X}^m(k)$  is at the end of a level  $n$  crossing)
  (This is always the case for  $n = m$ )
   $\kappa(m, n, k+1) = \kappa(m, n, k) + 1$ 
  If  $S_{\kappa(m,n,k)}^n = Z_{\kappa(m,n+1,k)}^{n+1}$  Then ( $\mathcal{X}^m(k)$  at end of level  $n+1$  crossing)
    Increment  $\mathcal{X}^{m,n+1}(k)$ 
     $S_{\kappa(m,n,k+1)}^n = 1$ 
    Generate  $Z_{\kappa(m,n+1,k+1)}^{n+1}$  using distribution  $p$ 
  Else ( $\mathcal{X}^m(k)$  not at end of level  $n+1$  crossing)
     $\mathcal{X}^{m,q}(k+1) = \mathcal{X}^{m,q}(k)$  for all  $q \geq n+1$  (†)
     $S_{\kappa(m,n,k+1)}^n = S_{\kappa(m,n,k)}^n + 1$ 
     $Z_{\kappa(m,n+1,k+1)}^{n+1} = Z_{\kappa(m,n+1,k)}^{n+1}$ 
  End If
  (Now determine the type of the new level  $n$  crossing)
  If  $S_{\kappa(m,n,k+1)}^n = Z_{\kappa(m,n+1,k+1)}^{n+1}$  Then (‡)
    If  $\alpha_{\kappa(m,n,k+1)}^{n+1} = i^+$  Then
       $\alpha_{\kappa(m,n,k+1)}^n = 1^+$ 
    Else
       $\alpha_{\kappa(m,n,k+1)}^n = -1^-$ 
    End If
  Else If  $S_{\kappa(m,n,k+1)}^n = Z_{\kappa(m,n+1,k+1)}^{n+1} - 1$  Then
    If  $\alpha_{\kappa(m,n,k+1)}^{n+1} = i^+$  Then
       $\alpha_{\kappa(m,n,k+1)}^n = 0^+$ 
    Else
       $\alpha_{\kappa(m,n,k+1)}^n = 0^-$ 
    End If
  Else If  $S_{\kappa(m,n,k+1)}^n$  is odd Then
     $\alpha_{\kappa(m,n,k+1)}^n = 0^+$  or  $0^-$  with equal probability
  Else
    If  $\alpha_{\kappa(m,n,k)}^n = 0^+$  Then  $\alpha_{\kappa(m,n,k+1)}^n = 1^-$  End If
    If  $\alpha_{\kappa(m,n,k)}^n = 0^-$  Then  $\alpha_{\kappa(m,n,k+1)}^n = -1^+$  End If
  End If
End Procedure

```

To update  $\mathcal{X}^m(k)$  to  $\mathcal{X}^m(k+1)$  we apply procedure **Increment** to  $\mathcal{X}^{m,m}(k)$ . **Increment** is recursively applied to all  $\mathcal{X}^{m,n}(k)$  such that  $S_{\kappa(m,q,k)}^q = Z_{\kappa(m,q+1,k)}^{q+1}$  for all  $m \leq q < n$ . For all  $n$  larger than this we get  $\mathcal{X}^{m,n}(k+1) = \mathcal{X}^{m,n}(k)$ .

Procedure **Increment** will always terminate after a finite number of recursive calls, provided we do not have  $S_{\kappa(m,q,k)}^q = Z_{\kappa(m,q+1,k)}^{q+1}$  for all  $q \geq m$ . However, if this is the case then for all  $n \geq m$  we put  $S_{\kappa(m,n,k+1)}^n = 1$ , generate  $Z_{\kappa(m,n+1,k+1)}^{n+1}$  according to the distribution  $p$  (independently of each other), and then generate

types consistently.

Since **Increment** only requires  $\mathcal{X}^m(k)$  to generate  $\mathcal{X}^{m+1}(k)$ ,  $\mathcal{X}^m$  is a Markov process.  $\square$

Clearly, given  $\mathcal{X}^{m,m}(k)$  we know  $X^m(k+1) - X^m(k)$ . So if we can simulate  $\mathcal{X}^{m,m}$  then we can simulate  $X^m$ .

Note that  $\mathcal{X}^m$  has a countable state space. Clearly it is transient if we include  $\kappa(m, n, k)$  in the crossing state, as  $\kappa(m, n, k)$  is a non-decreasing function of  $k$ . However, the chain is still transient if we remove the  $\kappa(m, n, k)$ . To see this, suppose we write  $\{s_m, s_{m+1}, s_{m+2}, \dots\} < \{t_m, t_{m+1}, t_{m+2}, \dots\}$  if  $s_n = t_n$  for all  $n > n_0$  and  $s_{n_0} < t_{n_0}$ , for some  $n_0 < \infty$ . This defines a partial ordering on  $\mathbb{Z}_+^{\mathbb{Z}_+}$ . We have  $\{S_{\kappa(m,m,k)}^m, S_{\kappa(m,m+1,k)}^{m+1}, \dots\} < \{S_{\kappa(m,m,k+1)}^m, S_{\kappa(m,m+1,k+1)}^{m+1}, \dots\}$  for all  $k$ . Thus, not only is  $\mathcal{X}^m$  transient, every state is visited at most once.

### 3 Simulation algorithm

The procedure **Increment** used to prove Theorem 2 is used as the basis of an algorithm to generate  $X^m$ . As noted previously, we can generate  $X^m$  from  $\mathcal{X}^{m,m}$ . The crux of our algorithm is the observation that we do not need all of  $\mathcal{X}^m(k)$  to generate  $\mathcal{X}^{m,m}(k+1)$ . Instead we can use a truncated version  $\{\mathcal{X}^{m,m}(k), \dots, \mathcal{X}^{m,n}(k)\}$ , where  $n = O(\log k)$ . This is because we only need to know  $\mathcal{X}^{m,n+1}(k)$  the first time there is a new level  $n+1$  crossing.

One way of achieving this is to put  $S^{m,n}(1) = S_1^n = 1$  for all  $n$ . In this case, if the first level  $n$  crossing ends at step  $k$  then  $\mathcal{X}^{m,n}(k) = (1, 1, Z_1^{n+1}, \alpha_1^n)$  where the  $Z_1^{n+1}$  are i.i.d. with distribution  $p$ . If  $\alpha^{m,n-1}(k) = 1^+$  then  $\alpha^{m,n}(k) = \alpha_1^n = 0^+$ . If  $\alpha^{m,n-1}(k) = -1^-$  then  $\alpha^{m,n}(k) = \alpha_1^n = 0^-$ .

While this provides a method of generating  $\mathcal{X}^{m,n}(k)$  when it is first required, in practice this approach is undesirable. The correlation structure of  $X^m$  is determined by the branching structure of the crossing tree. Restricting  $\mathcal{X}^m(1)$  as above effectively means we are conditioning  $X^m$  in some manner. As  $\mathcal{X}^m$  is transient, it does not have an equilibrium distribution, so we can not choose  $\mathcal{X}^m(1)$  in equilibrium. None-the-less, we can still choose  $\mathcal{X}^{m,n}(k)$  in a random fashion. The question we need to ask is: "for fixed  $n$ , if we observe an EBP process at a 'random' point in time  $k$ , what is the distribution of  $\mathcal{X}^{m,n}(k)$ ?"

Suppose that we have a sequence of i.i.d. non-negative random variables arranged in to families,  $X_{1,1}, X_{1,2}, \dots, X_{1,N(1)}, X_{2,1}, \dots, X_{k,N(k)}, \dots$ , where  $P(X_{i,j} \leq x) = F(x)$  and  $P(N(i) = n) = p(n)$ . Partition  $[0, \infty)$  into adjacent intervals  $[a, b)$  with lengths  $X_{i,j}$ , ordered as above. Then choose a point  $x$  'uniformly' in  $[0, \infty)$  and consider the size of the interval and family that contain  $x$ . Here when we say  $x$  chosen uniformly in  $[0, \infty)$  we mean in the limit as  $T \rightarrow \infty$  for  $x$  chosen uniformly in  $[0, T)$ .

We think of the  $X_{i,j}$  as level  $n$  crossing times and  $\sum_{j=1}^{N(i)} X_{i,j}$  as level  $n+1$  crossing times.

**Lemma 3** *Let  $\mathcal{P}$  be the partition above, let  $X^*$  and  $N^*$  be the interval length and family size of a 'uniformly' chosen point, and let  $J^*$  be the position of the chosen interval within its family. If  $\mu = \sum_x x p(x)$  and  $m = EX_{i,j}$  are finite then with*

probability 1, for  $1 \leq l \leq n$

$$P(N^* = n, J^* = l, x \leq X^* < x + dx | \mathcal{P}) = \frac{np(n)}{\mu} \frac{1}{n} \frac{1 - F(x)}{m} dx.$$

The proof is given in Section 4

At important consequence of this result is that  $N^*$  and  $J^*$  are independent of  $X^*$ . We also have that  $X^*$  is continuous even if the  $X_{i,j}$  are not, though this is not important here.

**Corollary 4** Let  $X$  be an EBP process. Observe  $X$  at some time  $t$  'uniformly' distributed over  $[0, \infty)$ , and let  $k$  be such that  $X(t) \in C_k^n$ . Then  $Z_{\kappa(n,n+1,k)}^{n+1}$  has distribution  $x p(x)/\mu$  and  $S_k^n$  is uniformly distributed over  $\{1, \dots, Z_{\kappa(n,n+1,k)}^{n+1}\}$ .

If  $S_k^n = Z_{\kappa(n,n+1,k)}^{n+1}$  then  $\alpha_k^n = 1^+$  or  $-1^-$  with equal probability. Otherwise if  $S_k^n$  is odd then  $\alpha_k^n = 0^+$  or  $0^-$  with equal probability. If  $S_k^n$  is even,  $S_k^n \neq Z_{\kappa(n,n+1,k)}^{n+1}$ , then  $\alpha_k^n = 1^-$  or  $-1^+$  with equal probability.

Finally, the sampling distributions of  $S_k^n$  and  $Z_{\kappa(n,n+1,k)}^{n+1}$  are independent of the length of  $C_k^n$ .

Our simulation algorithm uses a modified crossing state. For some  $n_{\max} = n_{\max}(k)$

$$\bar{\mathcal{X}}^m(k) = \{\mathcal{X}^{m,m}(k), \dots, \mathcal{X}^{m,n_{\max}}(k)\}.$$

We give a procedure **Expand** to increase  $n_{\max}$  when necessary. Let  $\bar{p}$  be the distribution given by  $\bar{p}(n) = np(n)/\mu$ .

```

Procedure Expand  $\bar{\mathcal{X}}^m(k)$ 
  While  $S^{m,n_{\max}}(k) = Z^{m,n_{\max}+1}(k)$  Do
     $\kappa(m, n_{\max} + 1, k) = 1$ 
    Generate  $Z^{m,n_{\max}+2}(k)$  using distribution  $\bar{p}$ 
    Generate  $S^{m,n_{\max}+1}(k) \sim U\{1, \dots, Z^{m,n_{\max}+2}(k)\}$ 
    If  $\alpha^{m,n_{\max}}(k) = 1^+$  Then
      If  $S^{m,n_{\max}+1}(k) = Z^{m,n_{\max}+2}(k)$  Then
         $\alpha^{m,n_{\max}+1}(k) = 1^+$ 
      Else If  $S^{m,n_{\max}+1}(k)$  is odd Then
         $\alpha^{m,n_{\max}+1}(k) = 0^+$ 
      Else
         $\alpha^{m,n_{\max}+1}(k) = -1^+$ 
      End If
    Else  $(\alpha^{m,n_{\max}}(k) = -1^-)$ 
      If  $S^{m,n_{\max}+1}(k) = Z^{m,n_{\max}+2}(k)$  Then
         $\alpha^{m,n_{\max}+1}(k) = -1^-$ 
      Else If  $S^{m,n_{\max}+1}(k)$  is odd Then
         $\alpha^{m,n_{\max}+1}(k) = 0^-$ 
      Else
         $\alpha^{m,n_{\max}+1}(k) = 1^-$ 
      End If
    End If
  End If

```

```

     $n_{\max} = n_{\max} + 1$ 
End While
End Procedure

```

We will also use procedure **Increment** with two changes. Firstly the line (†) is changed to

$$\mathcal{X}^{m,q}(k+1) = \mathcal{X}^{m,q}(k) \text{ for } q = n+1, \dots, n_{\max}$$

Secondly, at the point (†) we assign types differently if  $n = n_{\max}$ , as in this case we can not use  $\alpha^{m,n+1}(k+1)$  to determine  $\alpha^{m,n}(k+1)$ . Instead we determine  $\alpha^{m,n}(k+1)$  randomly.

```

If  $S_{\kappa(m,n,k+1)}^n = Z_{\kappa(m,n+1,k+1)}^{n+1}$  Then
  If  $\alpha_{\kappa(m,n,k)}^n = 0^+$  Then
     $\alpha_{\kappa(m,n,k+1)}^n = 1^+$ 
  Else  $(\alpha_{\kappa(m,n,k)}^n = 0^-)$ 
     $\alpha_{\kappa(m,n,k+1)}^n = -1^-$ 
  End If
Else If  $S_{\kappa(m,n,k+1)}^n$  is odd Then
   $\alpha_{\kappa(m,n,k+1)}^n = 0^+$  or  $0^-$  with equal probability
Else
  If  $\alpha_{\kappa(m,n,k)}^n = 0^+$  Then  $\alpha_{\kappa(m,n,k+1)}^n = 1^-$  End If
  If  $\alpha_{\kappa(m,n,k)}^n = 0^-$  Then  $\alpha_{\kappa(m,n,k+1)}^n = -1^+$  End If
End If

```

We can now give our simulation algorithm.

```

Procedure Simulate
  (Given  $\bar{\mathcal{X}}^m(k)$  and  $X^m(k+1)$  returns  $\bar{\mathcal{X}}^m(k+1)$  and  $X^m(k+2)$ )
  Expand  $\bar{\mathcal{X}}^m(k)$ 
  Increment  $\mathcal{X}^{m,m}(k)$ 
  If  $\alpha^{m,m}(k+1) = i^+$  Then
     $X^m(k+2) = X^m(k+1) + 2^m$ 
  Else
     $X^m(k+2) = X^m(k+1) - 2^m$ 
  End If
End Procedure

```

To initialise the crossing state we have the following

```

Procedure Initialise  $\bar{\mathcal{X}}^m$ 
   $n_{\max} = m$ 
   $\kappa(m, m, 1) = 1$ 
  Generate  $Z^{m,m+1}(1)$  using distribution  $\bar{p}$ 
  Generate  $S^{m,m}(1) \sim U\{1, \dots, Z^{m,m+1}(1)\}$ 
  If  $S^{m,m}(1) = Z^{m,m+1}(1)$  Then
     $\alpha^{m,m}(1) = 1^+$  or  $-1^-$  with equal probability
  Else If  $S^{m,m}(1)$  is odd Then
     $\alpha^{m,m}(1) = 0^+$  or  $0^-$  with equal probability

```

```

Else
     $\alpha^{m,m}(1) = 1^-$  or  $-1^+$  with equal probability
End If
End Procedure

```

We assume we are given  $X^m(1)$  (recall we take  $k = 1$  as our starting time). Given  $\bar{X}^m(1)$  and  $X^m(1)$  we get  $X^m(2) = X^m(1) \pm 2^m$  according to  $\alpha^{m,m}(1) = i^\pm$ .

### 3.1 Performance

On average,  $X^m$  starts a new level  $n$  crossing every  $\mu^{n-m}$  steps. It follows immediately that  $n_{\max}(k) = O(\log k)$ . To generate a new step of  $X^m$  it is only necessary to store the previous value and  $\bar{X}^m$ . Thus to generate  $n$  steps we require  $O(\log n)$  storage.

The expected operation count of procedure **Expand** is finite and independent of  $n_{\max}(k)$ . The operation count of procedure **Increment** is proportional to  $n_{\max}(k)$ . Thus the number of operations used by **Simulate** to generate  $n$  steps of  $X^m$  is of order

$$\sum_{k=1}^n \log k = \log n! \approx \log(\sqrt{2\pi n} e^{-n} n^n) = O(n \log n).$$

(Using Stirling's formula for the approximation.)

### 3.2 Crossing times

It is easily seen that the level  $m$  crossing times of the EBP process  $X$  are independent and have the same distribution as the normed limit  $W$  of the Galton-Watson branching process with offspring distribution  $p$  (up to some constant scaling). Thus, to simulate  $X$  at spatial resolution  $2^m$  we simulate  $X^m$  and use i.i.d. level  $m$  crossing times for the times between jumps. It is possible to sample from the distribution of  $W$  approximately, by simulating a finite number of generations and normalising by the expected population size.

In practice if  $m$  is small, then a rough approximation to the distribution of  $W$  is sufficient (even a constant approximation is  $m$  is small enough). It is possible to sample from the distribution of  $W$  with high accuracy very efficiently, but as this is really ancillary to the principal content of this paper we will not consider this problem further here.

### 3.3 Examples

We illustrate the algorithm with some simulated traces of EBP process. In Figure 2 we use the offspring distribution  $p(2k) = a(1-a)^{k-1}$  for  $a \in (0, 1)$ . This gives  $\mu = 2/a$  and  $H = \log 2 / (\log 2 - \log a)$ . From top to bottom we have  $(a, H) = (0.2, 0.3010)$ ,  $(0.5, 0.5)$  and  $(0.8, 0.7565)$ .

In Figure 3 we have four processes with the same  $H$  value of 0.5. The offspring

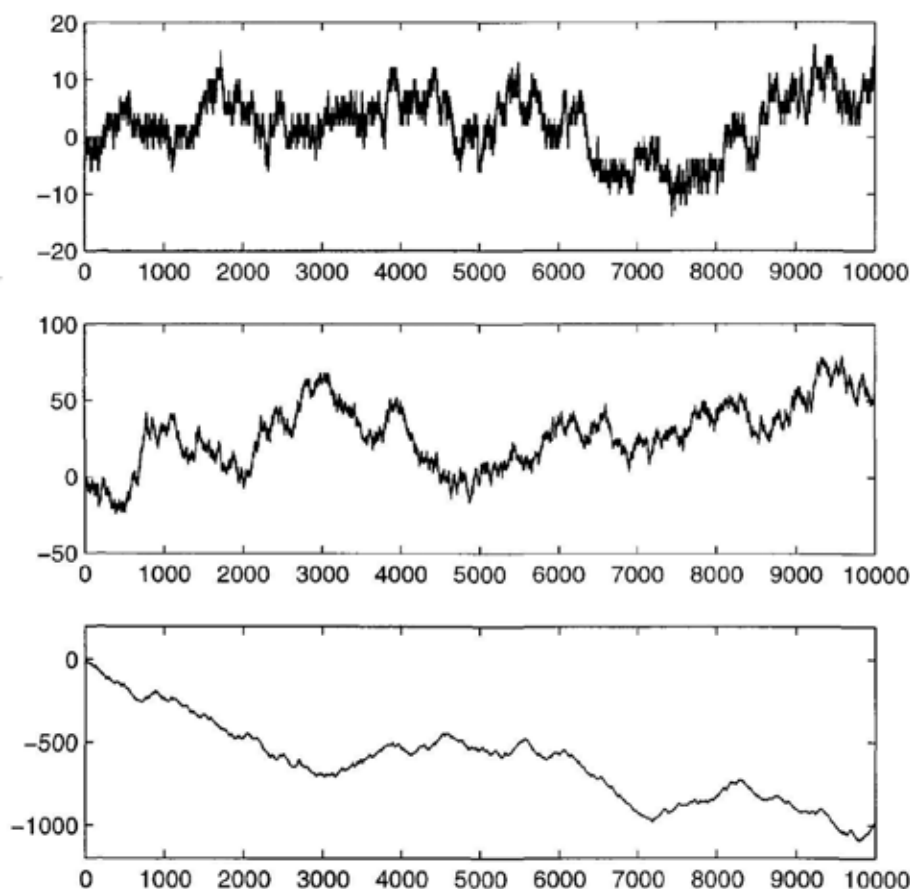


Figure 2. Self-similar traces with  $H = 0.3010, 0.5$  and  $0.7565$  respectively.

distributions for each process are in order

- Top left:  $p(2) = 0.5, p(6) = 0.5$ ;
- Top right:  $p(2) = 0.75, p(10) = 0.25$ ;
- Bottom left:  $p(2) = 0.9, p(22) = 0.1$ ;
- Bottom right:  $p(2) = 0.95, p(42) = 0.05$ .

## 4 Proofs

**Proof of Theorem 1** We construct a crossing of size 1. Given the self-similarity of the process, this can be scaled to give a sample of arbitrary length. The method we use dates back to Knight <sup>6</sup> and Barlow & Perkins <sup>7</sup>.

We first define a number of ancillary processes. For  $m \leq 0$  let  $Y^m$  be a random walk with steps of size  $2^m$  at times  $\mu^m \mathbb{Z}_+$ . Put  $Y^0(0) = 0$  and  $Y^0(1) = 1$ , then

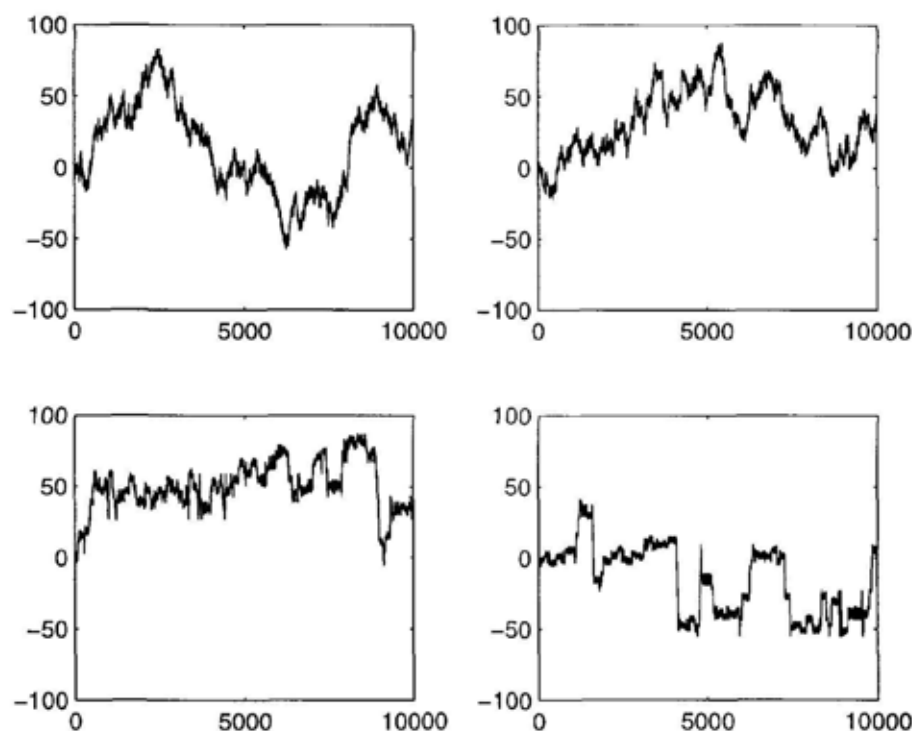


Figure 3. Self-similar traces with  $H = 0.5$  but different offspring distributions.

construct  $Y^{m-1}$  from  $Y^m$  by replacing step  $k$  of  $Y^m$  by a sequence of  $Z_k^m$  steps of size  $2^{m-1}$ , where  $P(Z_k^m = x) = p(x)$ . These are the level  $m-1$  sub-crossings of crossing  $k$  at level  $m$ . The  $Z_k^m$  are independent and identically distributed.

Each crossing can be up or down. A sequence of  $Z_k^m$  sub-crossings consists of  $(Z_k^m - 2)/2$  excursions followed by a direct crossing. An excursion is an up-down or a down-up pair; a direct crossing is an up-up or a down-down pair. If the parent crossing is up, then the sub-crossings end up-up, otherwise they end down-down. We allow each excursion to be up-down or down-up with equal probability (though note that other choices are possible).

We extend  $Y^m$  from  $\mu^m \mathbb{Z}_+ \rightarrow 2^m \mathbb{Z}$  to  $\mathbb{R}_+ \rightarrow \mathbb{R}$  by linear interpolation. Also let  $T^m = \inf\{t : Y^m(t) = 1\}$ , and for  $t > T^m$  put  $Y^m(t) = 1$ . The interpolated  $Y^m$  has continuous sample paths. We will show that with probability 1, as  $m \rightarrow -\infty$  the sample path of  $Y^m$  converges uniformly on any finite interval. The limiting sample path is thus continuous.

For  $n \leq m$  let  $T_0^{m,n} = 0$  and  $T_{k+1}^{m,n} = \inf\{t > T_k^{m,n} : Y^n(t) \in 2^m \mathbb{Z}, Y^n(t) \neq Y^n(T_k^{m,n})\}$ . If  $Y^n(T_k^{m,n}) = 1$  then we put  $T_{k+1}^{m,n} = \infty$ . The  $T_k^{m,n}$  are the level  $m$  hitting times of  $Y^n$ . The  $k$ -th level  $m$  crossing time of  $Y^n$  is  $W_k^{m,n} = T_k^{m,n} - T_{k-1}^{m,n}$ . For each  $m$  and  $k$ ,  $\{\mu^{m-n} W_k^{m,n}\}_{n=m}^{-\infty}$  is a Galton-Watson branching process. Thus since  $p$  is regular there exist i.i.d. continuous non-negative r.v.s  $W_k^m$  with mean 1

such that (see for example Athreya & Ney<sup>8</sup>)

$$W_k^{m,n} \rightarrow W_k^m \text{ with probability 1.}$$

Let  $T_k^m = \sum_{j=1}^k W_k^m = \lim_{n \rightarrow \infty} T_k^{m,n}$ .

Fix  $\epsilon > 0$  and  $T > 0$ . We will find a  $u$  such that for all  $r, s \leq u$  and  $t \in [0, T]$

$$|Y^r(t) - Y^s(t)| \leq \epsilon \text{ a.s.} \quad (2)$$

Given  $t \in [0, T]$ , let  $k = k(n, t)$  be such that

$$T_{k-1}^n \leq t < T_k^n$$

then for any  $r, s \leq n$

$$\begin{aligned} |Y^r(t) - Y^s(t)| &\leq |Y^r(t) - Y^r(T_k^{n,r})| + |Y^r(T_k^{n,r}) - Y^s(T_k^{n,s})| + |Y^s(T_k^{n,s}) - Y^s(t)| \\ &= |Y^r(t) - Y^r(T_k^{n,r})| + |Y^s(T_k^{n,s}) - Y^s(t)| \end{aligned} \quad (3)$$

noting that  $Y^r(T_k^{n,r}) = Y^s(T_k^{n,s}) = Y^n(k\mu^n)$ . Now, let  $j = j(u, T)$  be the smallest  $j$  such that

$$T_j^{n,u} > T$$

then as  $u \rightarrow \infty$ ,  $j(u, T) \rightarrow j(T) < \infty$  a.s., so we can choose a  $u$  such that for all  $q \leq u$

$$\max_{i \leq j} \{|T_i^{n,q} - T_i^n|\} < \min_{i \leq j} W_i^n \text{ with probability 1.}$$

Thus for any  $q \leq p$

$$T_{k-2}^{n,q} < t < T_{k+1}^{n,q}$$

and

$$|Y^q(t) - Y^q(T_k^{n,q})| = |Y^q(t) - Y^n(k\mu^n)| \leq 3 \cdot 2^n$$

since  $Y^q(T_{k-2}^{n,q}) = Y^n((k-2)\mu^n)$ ,  $Y^q(T_{k+1}^{n,q}) = Y^n((k+1)\mu^n)$  and in three steps  $Y^n$  can move at most distance  $3 \cdot 2^n$ . Applying this to (3) proves (2), taking  $n$  small enough that  $6 \cdot 2^n \leq \epsilon$ . Thus as  $\epsilon$  is arbitrary,  $Y^n$  converges to some (necessarily continuous)  $Y$  uniformly on all closed intervals  $[0, T]$ , with probability 1.

By construction  $Y(T_k^m) = Y^m(k\mu^m)$  for all  $m$  and  $k$ . The self-similarity (1) is also a direct consequence of the construction.  $\square$

**Proof of Lemma 3** Let  $T_k = \sum_{i=1}^k \sum_{j=1}^{N(i)} X_{i,j}$  and let  $\mathcal{P}_k$  be the partition of  $[0, T_k]$  given by  $X_{1,1}, \dots, X_{k,N(k)}$ . Given  $\mathcal{P}_k$ , choose  $x$  uniformly on  $[0, T_k]$  and let  $X_k^*$  and  $N_k^*$  be the interval length and family size of  $x$ . Let  $S_k(n) = \#\{i : 1 \leq i \leq k, N(i) = n\}$  then sending  $k \rightarrow \infty$

$$\begin{aligned} &P(N_k^* = n, J^* = l, X_k^* \leq x | \mathcal{P}) \\ &= \sum_{i=1}^k \sum_{j=1}^{N(i)} P(N_k^* = n, J^* = l, X_k^* \leq x | N_k^* = N(i), X_k^* = X_{i,j}, \mathcal{P}) \frac{X_{i,j}}{T_k} \end{aligned}$$



$$\begin{aligned}
&= \sum_{i=1}^k \sum_{j=1}^n I_{\{n\}}(N(i)) I_{\{l\}}(j) I_{[0,x]}(X_{i,j}) \frac{X_{i,j}}{T_k} \\
&= \frac{k}{T_k} \frac{S_k(n)}{k} \frac{1}{S_k(n)} \sum_{i=1}^k I_{\{n\}}(N(i)) X_{i,l} I_{[0,x]}(X_{i,l}) \\
&\rightarrow \frac{1}{\mu m} p(n) E X_{i,j} I_{[0,x]}(X_{i,j}) \text{ with probability } 1 \\
&= \frac{np(n)}{\mu} \frac{1}{n} \int_0^x \frac{1-F(y)}{m} dy.
\end{aligned}$$

Differentiating w.r.t.  $x$  gives the result.

By integrating/summing out the other terms, one can easily show that the marginal distributions of  $N^*$  and  $X^*$  are given by

$$\begin{aligned}
P(N^* = n) &= \frac{np(n)}{\mu}; \\
P(x \leq X^* < x + dx) &= \frac{1-F(x)}{m} dx.
\end{aligned}$$

Similarly, the conditional distribution of  $J^*$  given  $N^*$  is given by

$$P(J^* = l | N^* = n) = \frac{1}{n} \text{ for } 1 \leq l \leq n.$$

□

## References

1. J.R.M. Hosking. Modelling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20:1898–1908, 1984.
2. A.T.A. Wood and G. Chan. Simulation of stationary Gaussian processes in  $[0, 1]^d$ . *J. Computational Graphical Stat.*, 3:409–432, 1994.
3. V. Paxson. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *Comp. Comm. Rev.*, 27:5–18, 1997.
4. D.R. Cox. Long range dependence: A review. In *Statistics: An Appraisal*, David, H.A. & David, H.T., eds. Iowa State University Press, 1984.
5. O.D. Jones and Y. Shen. Estimating the Hurst index of a self-similar process via the crossing tree. To appear in *Signal Processing Letters*.
6. F.B. Knight. *Essentials of Brownian Motion and Diffusion*. Number 18 in AMS Math. Surveys. 1981.
7. M.T. Barlow and E.A. Perkins. Brownian motion on the Sierpinski gasket. *Prob. Th. Rel. Fields*, 79:543–623, 1988.
8. K.B. Athreya and P.E. Ney. *Branching Processes*. Springer-Verlag, 1972.

# **FRACTAL GEOMETRY IN THE ARTS: AN OVERVIEW ACROSS THE DIFFERENT CULTURES**

NICOLETTA SALA

*Accademia di Architettura of Mendrisio, Università della Svizzera italiana*

*Largo Bernasconi CH- 6850 Mendrisio, Switzerland*

*E-mail: nsala @ arch.unisi.ch*

Fractal, in mathematics, is a geometric shape that is complex and detailed in structure at any level of magnification. The word "fractal" was coined less than thirty years ago by one of history's most creative and mathematicians, Benoit Mandelbrot, whose work, *The Fractal Geometry of Nature*, first introduced and explained concepts underlying this new vision of the geometry. Although other mathematical thinkers like Georg Cantor (1845-1918), Felix Hausdorff (1868-1942), Gaston Julia (1893-1978), Helge von Koch (1870-1924), Giuseppe Peano (1858-1932), Lewis Richardson (1891-1953), Waclaw Sierpinski (1882-1969) and others had attained isolated insights of fractal understanding, such ideas were largely ignored until Mandelbrot's genius forged them at a single blow into a gorgeously coherent and fascinating discipline. Fractal geometry is applied in different field now: engineering, physics, chemistry, biology, and architecture. The aim of this paper is to introduce an approach where the arts are analysed using a fractal point of view.

## **1 Introduction**

Fractal geometry is a modern mathematical theory that radically departs from traditional Euclidean geometry. It describes objects that are scale symmetric, or self-similar. This means that when such objects are magnified, their parts are seen as an exact resemblance to the whole, the property continues with the parts of the parts and so on to infinity. These shapes are called fractals, and they must maintain a rough, jagged quality at every scale at which an object can be examined.

The nature and the characteristics of fractals are reflected in the word itself, coined by the Polish-born French mathematician **Benoit B. Mandelbrot** (b. 1924) from the Latin verb *frangere*, "to break", and from the related adjective *fractus*, "fragmented and irregular" [6, 15, 17]. The acceptance of the word "fractal" was dated in 1975. When Mandelbrot presented the list of publications between 1951 and 1975, date when the French version of his book was published. The people were surprised by the variety of the studied fields: noise on telephone lines, linguistics, cosmology, economy, games theory, turbulence. The multiplicity of the fields of application has played a central role in the genesis of Mandelbrot's discovery.

The first and simplest fractal object is the Cantor bar (also named Cantor set, or Cantor dust, by the nineteenth century German mathematician Georg Cantor). It is probably the most ancient known fractal. The Cantor bar may be realized by dividing a line in 3 parts and removing the middle part. This procedure is iterative and it is repeated indefinitely, first on the 2 remaining parts, then on 4 parts produced by that operation, and so on, until the object has an infinitely large number of parts each of which is infinitely small.

In 1904, Koch has published the work on his famous curve [25]. Then came Sierpinski's triangle in 1916 [23]. Few twentieth century mathematicians noticed that there were more sophisticated means to define the dimension of an object. Fundamental work was done by Hausdorff (1919), then developed by Besicovitch (1935). The Hausdorff-Besicovitch dimension has played, later on, a major role in the domain of the fractal geometry.

Irregularity, self-similarity between the original structure and its smaller constitutive fragments, form invariance under changes of measure (scaling) and iteration of unit generator, are main properties which characterize the fractal objects.

Mandelbrot used the term "self-similar" for the first time in 1964, in an internal report at IBM, where he was doing research, and in the title of a 1965 paper. A fractal object is self-similar. This means that as viewers peer deeper into the fractal image, we can notice that the shapes seen at one scale are similar to the shapes seen in the detail at another scale. It is possible to demonstrate that the fractal shapes and the self-similarity were known to the artists in different cultures. The artists have always used Euclidean geometry in the one or other way for their works, but they have also used the self-similarity, although they may not have been conscious of that. For example, the "golden section", one of the most important proportion-rules, is generated using a procedure based on a fractal sequence. The art can be interpreted as a way for finding the basics of beauty and harmony that are found in the laws of Nature [4, 5]. In this way chaos and fractal geometry may help to explain and prove the "rules" of beauty.

## 2 Fractal components in the arts

As we shall see, fractal geometry appears in the arts for reasons other than mimicking patterns in Nature. Our fractal analysis in the arts can be divided in two steps:

- an unconscious use of the fractal components or fractal properties, for example the self-similarity [4, 5, 10], the bifurcation processes [7], the L-systems [18];
- a conscious use of fractal geometry, for example to break the symmetry, to mimic the chaotic shapes, or to realize electronic paintings using the computer graphics and fractal procedures [4, 5, 13, 18, 21, 22].

### 2.1. Unconscious fractal components in the arts

In different cultures and in different styles are present many unconscious fractal components [20]. An interesting example is the capital of an Egyptian temple column (figure 1a). Ancient Egyptian cosmogony, often used to represent the development of the universe the white lotus flower [15, 21]. The lotus' corolla is organized in petals within petals within petals, in this way the lotus represented the cosmos on smaller and smaller scales. This is a clear example of self-similarity. We can compare the stylized lotus petals and the similarity of this representation to the first few stages of a Cantor's bar (figure 1b).



Figure 1. This Egyptian capital a) shows an interesting analogy with a Cantor's bar b)

The images in figure 2 show other fractal components present in a Western capital and their comparison with a natural object and with a fractal object [22]. The self-similar shapes are also present in the Gothic and in the Baroque arts [13, 22, 24].

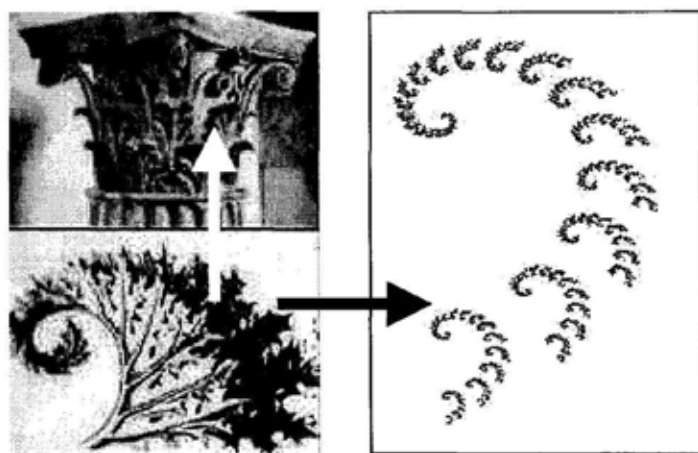


Figure 2. A Western capital, a natural object and a fractal object

In the book entitled *Geometry from Africa*, the mathematician Paulus Gerdes describes a enchanting guided tour of geometric ideas created by the people of the south of the Sahara, and encoded in woven designs, carved patterns, sand drawings, wooden models, and other products [12]. He describes a variety of geometrically decorated artefacts, from rock paintings and engravings to decorated pots and hand-woven materials, some of which are more than 2,000 years old. He also describes some interesting fractal components present in the African arts. In particular, a pyramidal basket is woven, and that is called *Eheleo* in the Makhuta language. It is used as a funnel in the product of salt. We can find it in the North of Mozambique, in the South of Tanzania, in the Congo/Zaire region and in Senegal [11]. The *Eheleo*, shown in the figure 3, has the shape of a triangular pyramid: the base is an equilateral triangle and the other three faces are isosceles right triangles [11, p. 83]. The figure 4 illustrates the composition of a structure that explores the right angles of the *Eheleo*, and it shows that the idea of self-similarity is known in the African art [11, p. 83]. Gerdes affirms: "The height of each new pyramid that is added to the structure is a fixed proportion of the last one (in figure this proportion is  $2/3$ ). Another way to produce a fractal architectural structure with *Eheleo* pyramids is by joining differently sized *eheleo* pyramids placed on their equilateral-triangular base" [11, p. 82].

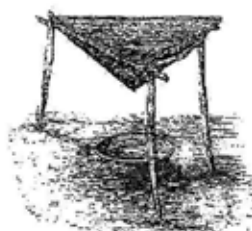


Figure 3. *Elheo*-funnel

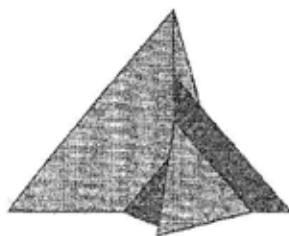


Figure 4. The composition of an *Elheo*

The figure 5a shows an African sculpture, the comparison to a fractal binary tree (figure 5b) is manifest [8]. We can find other fractal components in the Tuareg leatherworks, in the Mauritanian stoneworks, and in the Ghanaian sculptures realized as Sierpinski's tetrahedron [7, 8].

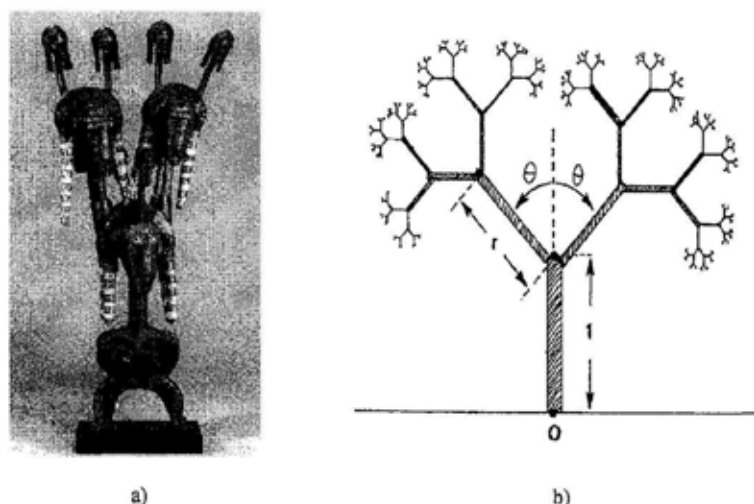


Figure 5. African sculpture a) and a fractal tree b)

In the Japanese arts we can find the presence of the self-similar spirals and of the process of bifurcation in the Hokusai's works. Hokusai's full name was **Katsushika Hokusai** (1760-1849), Japanese painter and wood engraver, born in Edo (now Tokyo). He is considered one of the outstanding figures of the Ukiyo-e, or "pictures of the floating world" (everyday life), school of printmaking. Hokusai entered in the studio of his countryman Katsukawa Shunsho in 1775 and there learned the new, popular technique of woodcut printmaking. Between 1796 and 1802 he produced a vast number of book illustrations and colour prints (perhaps as many as 30,000) that drew their inspiration from the traditions, legends, and lives of the Japanese people. Hokusai's most typical wood-block prints, silk-screens, and landscape paintings were done between 1830 and 1840. The curved lines characteristic of his style gradually developed into a series of spirals that imparted the utmost freedom and grace to his work, as in *Raiden, the Spirit of Thunder*. Figure 6 illustrates *The Breaking Wave Off Kanagawa*, also called *The Great Wave* (1831). Woodblock print from Hokusai's series *Thirty-six Views of Fuji*, which are the high point of Japanese prints. The original is at the Hakone Museum in Japan. In *The Great Wave*, there are three boats among the turbulent, broken waves. The boats mould into the shapes of the engulfing waves. Some humans are tossed around under giant waves, while the sacred, snow-capped Mount Fuji is a hill in the distance. Observing figure 6, we can note the presence of some different self-similar spirals. This fractal motif is present in others Japanese works on silk (as shown in figure 7). In *Amida falls* (1834-1835), shown in figure 8a, Hokusai represents the falls as a sub-harmonic function illustrates in figure 8b [10]. In the Hokusai's *Kirifuri Waterfall at Mount Kurokami in Shimotsuke province* (1832), Nelson-Atkins Museum of Art, Kansas City (Missouri), shown in figure 9, there is a realistic depiction of Kirifuri waterfall, one of the three famous waterfalls of Nikko. Particularly impressive is the analogy of a fractal process of bifurcation.

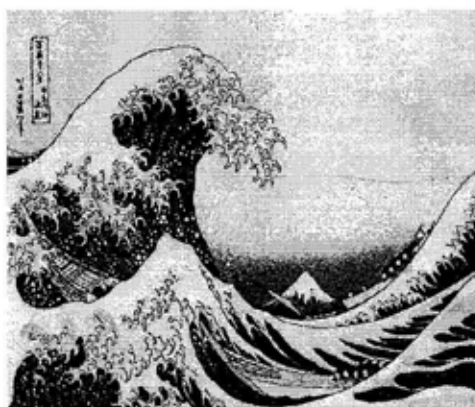
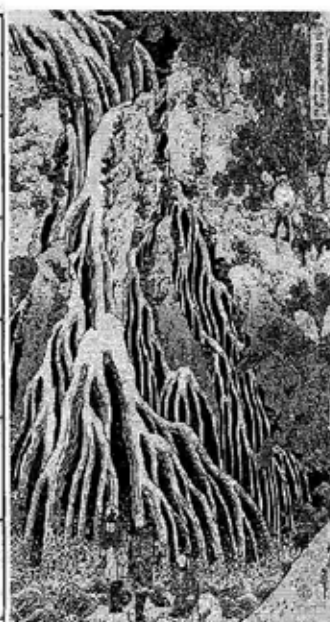
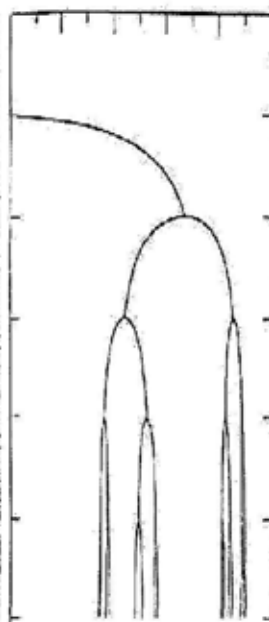


Figure 6. Hokusai's *The Breaking Wave Off Kanagawa*

Figure 7. Japanese self- similar spirals on the silk



a)

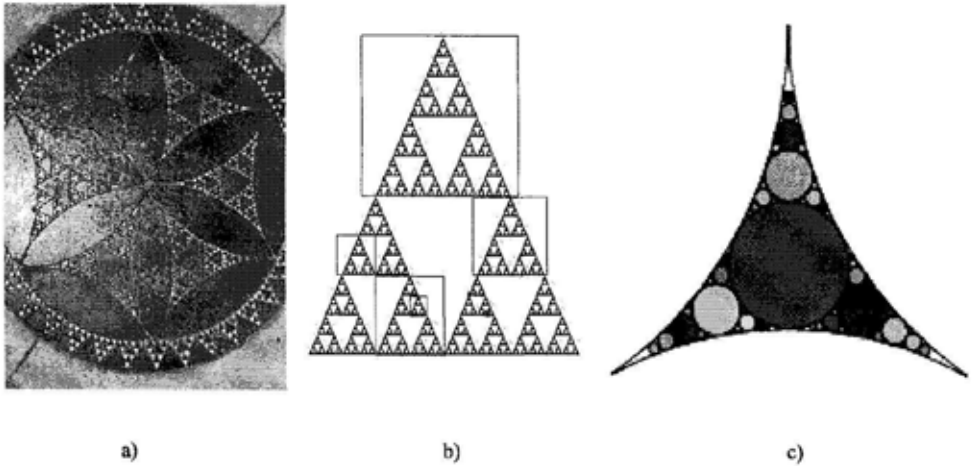
b)

Figure 8. Hokusai's *Amida Falls* (1834-1835) a) and the sub harmonic function b)

Figure 9. Hokusai's *Kirifuri Waterfall at Mount Kurokami in Shimotsuke province* (1832)

In the Western art we can find the oldest handmade fractal object in the Cathedral of Anagni (Italy). Inside the cathedral, built in the year 1104, there is a floor, illustrated in figure 10a, that is adorned with dozens of mosaics, each in the form of a Sierpinski fractal gasket (shown in the Figure 10b), but it impressive the analogy with an Apollonian gasket, shown in figure 10c [20]. The Apollonian gasket corresponds to a limit set that is

invariant under a Kleinian group [27, p. 986]. Kleinian group is a finitely generated discontinuous group of linear fractional transformations  $z \rightarrow (az + b)/(cz + d)$  acting on a domain in the complex plane [16].



**Figure 10.** The floor of the Cathedral of Anagni a) the Sierpinski gasket b) and the Apollonian gasket c)

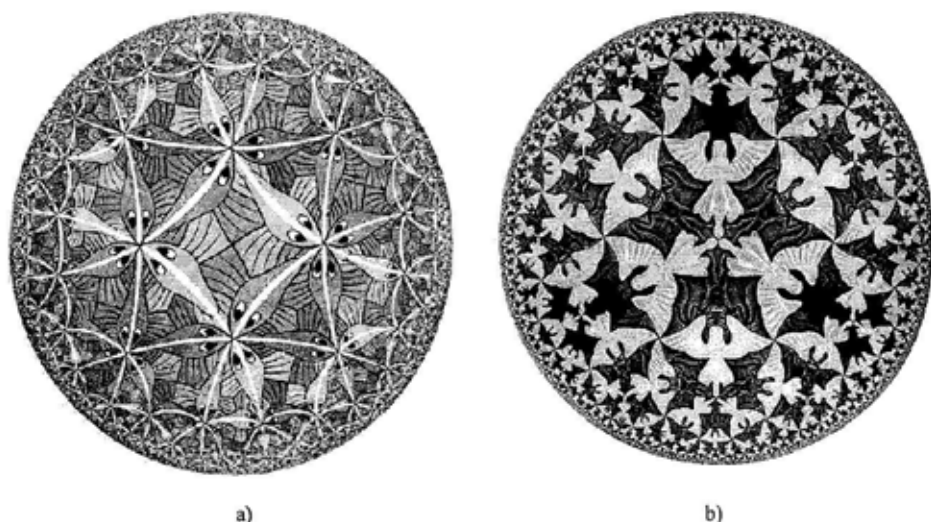
The Dutch graphic artist **Maurits Cornelis Escher** (1898-1972) was fascinated by the geometry, when he met the beauty of the fourteenth century Moorish palaces and in particular, by the decorative majolica tilings which adorned many of the surfaces of the palaces. One building which had an immense influence on the Escher's life was the *Alhambra Palace* in Granada (Spain). Unlike the Moors, Escher used, in his versions of the tilings, the objects created by his fantasy, for example snakes, chameleons, reptiles, birds, and ghosts. He realized a number of attempts using the division of the Euclidean plane, but he was unsatisfied about the poor quality of his final works, and he left regular division for a number of years. When the artist read the Pólya's 1924 paper on plane symmetry groups, he did not understand the abstract concepts of the groups described in the Pólya's work, but he understood the seventeen plane symmetry groups presented there. Between 1937 and 1941 Escher worked on possible periodic tilings producing forty-three coloured drawings dedicated to the symmetry types.

Over the years that followed Escher made numerous woodcuts utilising each of the seventeen symmetry groups. These extensive investigations culminated in 1941 with his first notebook entitled: *Regular Division of the plane with Asymmetric congruent Polygons*. In 1958 the artist met the British mathematician Harold Scott MacDonald Coxeter (1907-2003) and they became life-long friends. Escher read an article written by Coxeter, and again he was unable to understand the text, but he was able to determine the rules regarding hyperbolic tessellations, observing the diagrams in the paper. The Dutch artist found in the hyperbolic geometry the way to realize high quality works. For this reason to thank Coxeter, Escher sent to him a copy of his works *Circle Limit I* (1958), realized with the model of Poincaré presented in the Coxeter's paper. He produced many more prints with the hyperbolic geometry.

Escher used the fractal geometry and the self-similarity in unconscious way, in fact he did not mention them inside his engravings, but the property of the self-similarity is

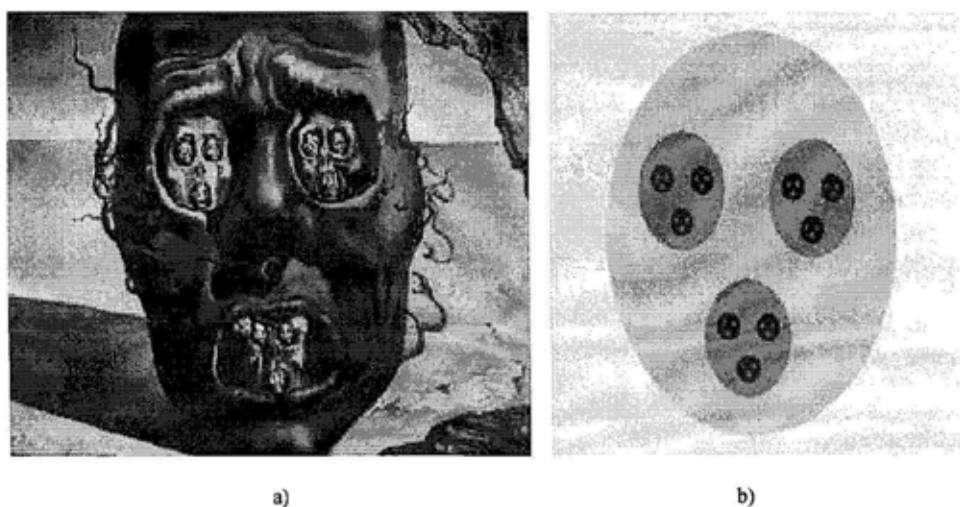


evident for example in the *Circle Limit III* (1959) (illustrates in figure 11a) and in the *Circle Limit IV (Heaven and Hell)*, 1960) (shown in figure 11b).



**Figure 11.** *Circle Limit III* a) and *Circle Limit IV* b) are two examples of the Escher's unconscious self-similarity

The Spanish surrealist painter, **Salvador Dali** (1904-1989) has applied some fractal components. His *Visage of War* (1940), oil on canvas, provides a clear example of self-similarity in the art. It shows a geometric representation like a "Russian doll" where the skulls are nested inside other skulls (figure 12a). The result of this kind of nesting is a shocking vision that emphasizes the drama of the war. If Dali's painting is seen using the fractal point of view, we can find a particular kind of fractal set (figure 12b) that corresponds to the Dali's work [21, 22].



**Figure 12.** Dali's *Visage of War* (1940) a), and the fractal set associated to the Dali's painting b)



In the Indian arts it is usual to find intricate works connected to the fractal geometry: the *kolams*. The *kolam* is a decorative draw that embellishes the courtyards and the doorstep of the homes in the villages of southern India. *Kolams* are known by different names in different parts of India. *Muggulu* in Andrapradesh, *Hase* in Karnataka, *Chowkpurna* in Uttar Pradesh, *Rangoli* in Gujarat and Maharashtra, and *Alpana* in Bengal and Assam. Perhaps five millennia old, this artefact is described in many ancient Sanskrit texts. The *kolam* can cover areas up to three meters by three meters. It consists of some small geometrical patterns repeated many times, that comprises lines, dots, squares, circles, triangles, lotus, shells, leaves, trees and flowers connected in very complicated ways. No gaps to be left anywhere between the line for evil spirits to enter. It is Hindu belief that the geometrical patterns and designs applied with rice flour at the entrance to a home, invites Goddess Lakshmi into the household, and drives away the evil spirits. *Kolam* is an auspicious symbol, and it is the most important kind of female artistic expression.

Prusinkiewicz and Hanan have shown that many of the more elaborate *kolams* can be generated using the L-systems, and they are fractal [18]. The figures 13a and 13b illustrate two examples of *kolams*: the *Anklets of Krishna* and the *Snake*. The *Anklets of Krishna* has defined by the axiom:  $-x--x$  and it has the following production rule:  $x \rightarrow xFx--xFx$  (with angle 45 degrees). The *Snake* has defined by the axiom:  $F+xF+F+xF$  and it has the following production rule:  $x \rightarrow xF-F-F+xF+F+xF-F-F+x$  (with angle 90 degrees).

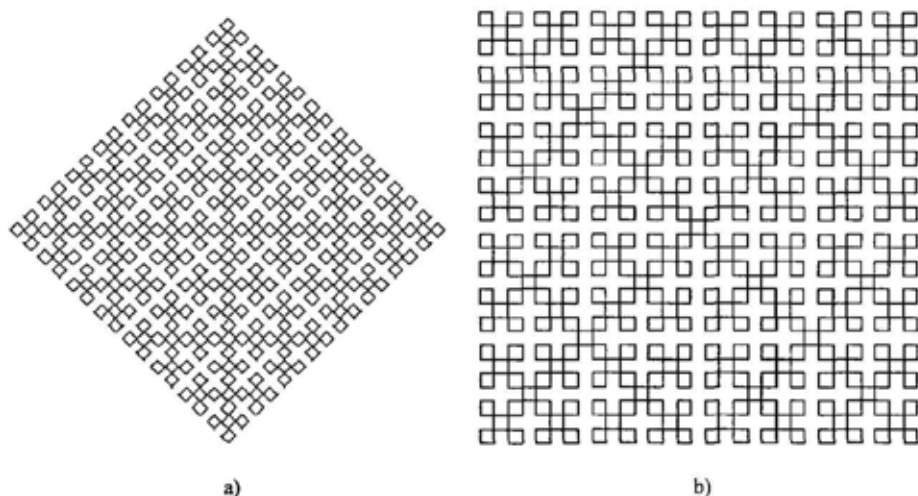


Figure 13. Two examples of *kolams*, the *Anklets of Krishna* a), and the *Snake* b)

In classical Islamic Art, ornamentation has a significant value that can be seen in every artistic expression from the carved marble panels of grand Mughal doorways in India, to the blue ceramic tiles of *Masjids* in Iran, to the elegant decorative artefacts in Syria. Arabesque, its style, composition and principles can be found in every objet d'art of classical Islam. The characteristic of Islamic art is a preference for covering surfaces with patterns composed of geometric or vegetal elements like flowers, foliage, and an extensive use of abstract geometric designs [1, 2, 14, 24, 26]. One can find the principles of geometry along with a keen sense of balance in composition strongly embedded in Islamic art. El-Said and Parman put forward a system in which geometrical grids are

broken down into identical units which are repeated in regular sequence [9]. There are three principle area of Islamic two-dimensional artistic expression: the calligraphy, the floral idioms, and the geometric patterns.

**Jay Bonner** has analyzed the three tradition self-similarity in fourteenth and fifteenth century Islamic geometric ornaments [3]. Bonner has classified three types of the self-similar Islamic geometric patterns. The first type is characterized by a primary repetitive geometric pattern, with a reduced scale on a secondary geometric pattern that has the same geometric characteristics as the primary, and it fills the complete background of the primary pattern. Bonner indicates these patterns as **Self-Similar Type A Patterns**, an example is shown in figure 14 [3, p.4]. The second type, called by Bonner: **Self-Similar Type B Patterns**, is realized on a primary geometric pattern, where the lines of which have been widened to a proportion that allows for a secondary geometric pattern, which has the same geometric characteristic of the primary pattern but at a reduced scale, to be placed within the widened lines (figure 15) [3, p.3]. The third type, called by Bonner: **Self-Similar Type C Patterns**, is present in Morocco and in Andalusia (Spain). The self-similar patterns of these regions are based on colour contrast to emphasize the primary design. An example of Self-Similar Type C Patterns is illustrated in figure 16 [3, p.4].

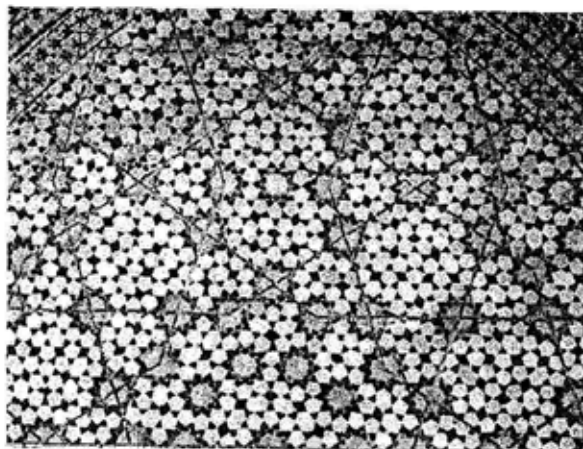


Figure 14. Self -Similar Type A design from the *Drab-i Imam* (Isfahan, Iran)

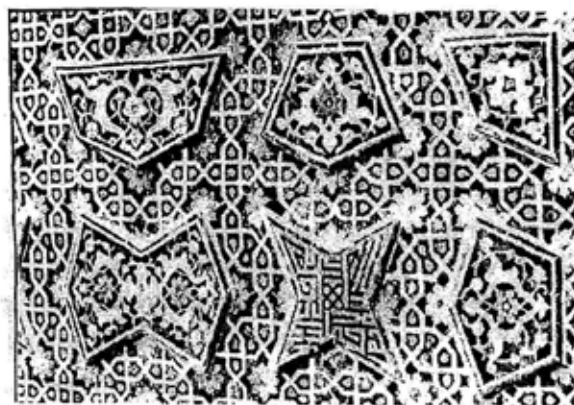


Figure 15. Self -Similar Type B design from the *Masjid-i Jami* (Isfahan, Iran)

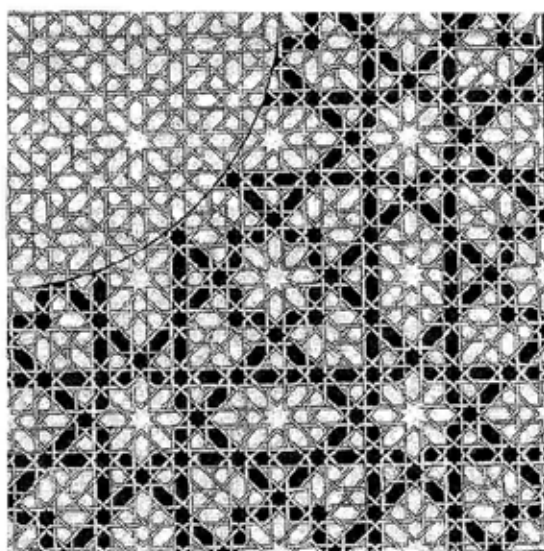


Figure 16. Self-Similar Type C design from the *Alcazar* (Seville, Spain)

## 2.2. Conscious fractal components in the arts

We can think about the Vincent Van Gogh's dense swirl of energy around the objects and the stars as a kind of chaotic shape, and the drip paintings realized by Jackson Pollock (1912-1956) as a kind of complexity. Van Gogh and Pollock used the fractal geometry in unconscious way, but now, at the beginning of the twenty-first century, there is a conscious use of the fractal geometry in the art. Art has become a self referential and self-reproducing system. The conscious use of the fractal components is a recent discovery by the twentieth century artists as a result of a specific and conscious act of creativity [4, 5]. Today the artists are excited by the recognition that the properties of the fractal geometry involve an aesthetic sense. Therefore, the rise of fractals has also democratized art and posed a serious question for contemporary artists.

The German painter and poet **Max Ernst** (1891-1976) was a member of the Dada movement, and after he became a surrealist. Ernst has applied the method of viscous fingering to realize his textured images that evoked dream-like worlds. This method was used by several artists, specifically with the technique named "decalcomania". Some examples of Ernst's works are: *Mythological figure-woman* (1940), *Europe After the Rain II* (1940-42), shown in figure 17, *Three well-tempered cypresses* (1949), and *Blue mountain and yellow sky* (1959). Ernst was a pioneer in the method named "frottage", in which a sheet of paper is placed on the surface of an object and then pencilled over until the texture of the surface is transferred. **Oscar Dominguez** (1906-1958) used this method with ink instead of paint in his *Decalcomanias* (1936), *The Lim-Bicycle* (1936).

In the late 1960s, **Larry Poons** (b. 1937) abandoned his simple "dot and blip" paintings to realize much more complex textured works that one critic of art described as "geological ... alluvial ... muddy" [4, 5]. Poons prepared a canvas by first affixing an undersurface of lightweight material (bits of foam toys, polyester fibres, and the like). Then he suspended the canvas vertically and flung buckets of paint on it. The artist experimented the viscosity and the colour of the paint, the drying time between impacts, the force and direction of the throw, as variables to realize the painting. What comes out,

is a collection of large blobs, surrounded by a halo of smaller blobs, themselves decorated with still smaller blobs, and so on. Poons used a kind of self-similarity, and Daniel Robbins described Poons' work as fractal [19].

The figure 18 shows a painting realized by the New York artist **Edward Berko**. Briggs refers on the Berko's idea of fractal and chaos: "to explore the manifestation of structure in nature. I paint in order to explore the potential of fractal geometry, to express a reinterpreted aesthetic of nature" [4, p. 170].

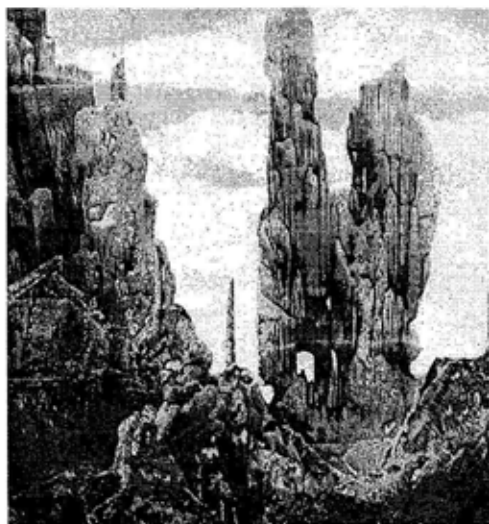


Figure 17. Ernst's *Europe After the Rain II* (1940-42)

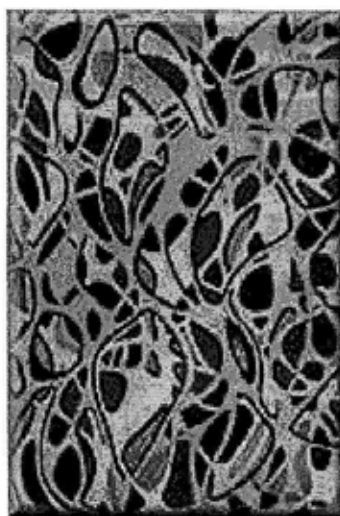


Figure 18. Berko's fractal painting

### 3 Conclusions

This paper has presented only some particular aspect of the fractal geometry in the arts. In particular, we have described the self-similarity, and the bifurcation processes in different cultures and through different periods. The self-similarity is present as an aesthetic property in all cultures, for example in African, Mesoamerican, Western, Japanese, Chinese, Hindu and Islamic cultures.

We can also apply the fractal geometry to realize the electronic paintings using the computer graphics. In this way, fractal geometry can generate new kinds of artists that use the monitor screen instead of the marble and the granite or the canvas [4, 5]. We are sure that the fractal geometry is helping to define a new aesthetic sense where the broken symmetry, the self-similarity and the bifurcation processes can play a central role. The fractal geometry and its connection between chaos and complexity theories can help to introduce the new complexity paradigm in the arts.

Briggs affirms: "When painters juxtapose multiple self-similar forms and colours on canvas, or composers transform a sequence of notes into multiple self-similar forms by varying the rhythm and projecting the sequence of notes into different sections of the orchestra, they create a tension that gives birth to lucid ambiguities. Such artistic juxtapositions might be called "reflectaphors" because the self-similar forms reflect each other yet contain, like metaphors, a tension composed of similarities and differences between the term. This reflectaphoric tension is so dynamic that it jars the brain into wonder, awe, perplexity, and a sense of unexpected truth or beauty" [4, p. 174].

#### 4 References

1. Blair S.S. and Bloom J.M., *The Art and Architecture of Islam 1250-1800* (Yale University Press, London, 1995).
2. Blair S.S. and Bloom J.M., *Islamic Art* (Phaidon Press, London, 1997).
3. Bonner J., Three Traditions of Self-Similarity in Fourteenth and Fifteenth Century Islamic Geometric Ornament, *Isama-Bridge 2003 Conference Proceedings*, Granada, Spain, pp. 1- 12.
4. Briggs J., *Fractals The Patterns of Chaos* (Thames & Hudson, London, 1992).
5. Brigg J., *Estetica del caos* (Red Edizioni, Como, 1993).
6. Crilly J., Earnshaw R. A. and Jones H., *Fractals and Chaos* (Springer - Verlag, New York, 1991).
7. Donato F. and Lucchi Basili, L., *L'ordine nascosto dell'organizzazione urbana* (Franco Angeli Editore, Milano, 1996).
8. Eglash R., *African Fractals* (Rutgers University Press, 1999).
9. El-Said I. and Parman A., *Geometric Concepts in Islamic Art* (Dale Seymour Publications, London, 1976).
10. Fivaz R., *L'ordre et la volupté* (Press Polytechniques Romandes, Lausanne, 1988).
11. Gerdes P., On some Geometrical and Architectural Ideas from African Art and Craft, Williams K. (edited by), *Nexus II: Architecture and Mathematics* (Edizioni Dell'Erba, Fucecchio, 1998) pp. 75 -86.
12. Gerdes P., *Geometry from Africa: Mathematical and Educational Explorations* (Mathematical Association of America Washington, D.C., 1999).
13. Hersey G.L., *The Monumental Impulse* (The Mit Press, Cambridge, Massachusetts, London, 1999).
14. Hillebrand R., *Islamic Art and Architecture* (Thames & Hudson, London, 1998).
15. Mandelbrot B., *The Fractal Geometry of Nature* (W.H. Freeman, New York, 1982).
16. Mumford D., Series C. and Wright D. J., *Indra's Pearls: An Atlas of Kleinian Groups* (Cambridge University Press, Cambridge, 2002).
17. Peitgen H-O. and Richter P.H., *The Beauty of Fractals: Images of Complex Dynamical Systems* (Springer-Verlag, Berlin, 1986).
18. Prusinkiewicz P. and Hanan J., *Lindenmayer Systems, Fractals, and Plants* (Springer-Verlag, New York, 1989).
19. Robbins D., *Larry Poons: Creation of the Complex Surface, Larry Poons: Paintings 1963-1990* (Salander-O'Reilly Galleries, New York, 1990).
20. Sala N., The presence of the Self-Similarity in Architecture: Some examples, Novak M. M. (ed.), *Emergent Nature* (World Scientific, Singapore, 2002) pp. 273- 283.
21. Sala N. and Cappellato G., *Viaggio matematico nell'arte e nell'architettura* (Franco Angeli, Milano, 2003).
22. Sala N. and Cappellato G., *Architettura della complessità. La geometria frattale tra arte, architettura e territorio* (Franco Angeli, Milano, 2004, in print).
23. Sierpinski W., Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée, *C. R. Acad. Paris* 162 (1916).pp. 629-632.
24. Stierlin H., *Islamic Art and Architecture: From Isfahan to the Taj Mahal* (Thames & Hudson, New York, 2002).
25. von Koch H., Sur une courbe continue sans tangente, obtenue par une construction géométrique élémentaire, *Arkiv för Matematik* 1 (1904) pp. 681-704.
26. Wilson E., *Islamic Designs for Artists and Craftspeople* (Dover Publications, New York, 1988).
27. Wolfram S., *A New Kind of Science* (Champaign, IL: Wolfram Media, 2002) p. 986.

# FRactal Properties and Characterization of Road Profiles

PIERRICK LEGRAND

*IRCCyN, 1 rue de la Nôé, 44321 Nantes, France  
e-mail: pierrick.legrand@irccyn.ec-nantes.fr*

JACQUES LÉVY VÉHEL

*IRCCyN, 1 rue de la Nôé, 44321 Nantes, France and  
Projet Fractales, INRIA Rocquencourt, 78153 Le Chesnay Cedex, France  
e-mail: vehel@irccyn.ec-nantes.fr*

MINH-TAN DO

*LCPC, route de Bouaye, 44340 Bouguenais, France*

A major problem in road engineering is to understand the mechanisms of friction between rubber and the road. Several authors have claimed that road profiles are fractal, and that this fractality is related to the friction properties of the road. We study road profiles obtained using tactile and laser captors. These profiles belong to different category characterized by different friction coefficients. We find that all our profiles indeed display strong fractal behaviour in terms of both correlation exponents and regularization dimension over a large range of scales. However, neither of these fractal parameters seem to be related to friction. We then use a local fractal parameter, namely the pointwise Hölder exponent. We show that this exponent does discriminate profiles which have different friction properties.

## 1 Introduction and background

An important problem in road engineering is to understand the mechanisms of friction between rubber and the road. This is a difficult problem, since friction depends on many parameters: The type of rubber, the type of road, the speed, . . .

Several authors have shown that most road profiles are fractal<sup>1,2,9</sup> on given ranges of scales. Such a property has obvious consequences on friction, some of which have been investigated for instance in<sup>1,4</sup>. The main idea is that, in the presence of fractal roads, all scales of irregularity contribute to friction<sup>3</sup>.

In this work, we verify that road profiles finely sampled using tactile and laser captors are indeed fractals. More precisely, we show that they have well-defined correlation exponent and regularization dimension over a wide range of scales. However, although we deal with various classes of profiles which have different friction coefficients, we find that such *global* fractal parameters are not able to discriminate between the profiles. This means that friction may have relatively low correlation with fractional dimensions or correlation exponents. We then compute a *local* parameter called the pointwise Hölder exponent. Our experiments show that this exponent allows to separate road profiles which have different friction coefficients.

## 2 The road profiles

Our profiles are provided by the LCPC (Laboratoire Central des Ponts et Chaussées). These profiles correspond to coatings with various gravel, and are



characterized by different friction coefficients. A major aim of our study is to be able to relate these friction coefficients to fractal features of the profiles.

The samples consists of rectangular plaques with size  $100 \times 150 \text{ mm}^2$ . Their surface is made of gravels cast into a a synthetic resin mould. The gravels come from thirteen different gravel pits, with a size varying between 6 and 10 mm.

The manufacturing of the plaques consists in arranging the gravel in a flat-bottomed rectangular mould, then filling the mould with fine sand and after that adding a quick setting resin. After removing from the mould, the plaques go through laboratory polishing cycles, that we describe briefly. A mixture of water and fine abrasives are thrown up to the surfaces with a 10 MPa pressure. This processing induces certain changes in the microtexture of the gravel: The gravel originating from little polishable rocks keeps its initial microtexture; the one coming from highly polishable rocks loses its initial microtexture and becomes very smooth. Laboratory polishing thus makes it possible to emphasize the difference in microtexture between the different samples.

1D profiles have been sampled on the plaques through three different procedures: one using a tactile captor, and two using a laser captor. We briefly discuss these.

#### Tactile measures

The details of the tactile sensor are as follows. The radius of the contact tip is 2 microns. The sensor's depth of field is 6mm. To avoid a potential locking of the contacting tip during its shifting on the tested surfaces, the gap in between the stones is filled with resin. Fifteen profiles are measured on each plaque within a zone of size approximately  $75 \times 125 \text{ mm}$ . The length of the profiles varies from 12 to 25mm according to the nature of the surfaces to be measured, reaching a total length of about 300mm altogether. The sampling step is 4 microns, and the samples contains approximately 3100 to 6000 points.

#### Laser captor

The laser acquisition system developed at LCPC, based on an Imagine Optics captor, allows to modes: One uses a locking of the height, as the other does not. These two modes will be referred to in the following as locked and non-locked (see details in <sup>3</sup>). Again, fifteen profiles are measured on each plaque within a zone of size approximately  $75 \times 125 \text{ mm}$ . The sampling step is 10 microns. The length of each profile is 125 mm, resulting in a sample size of 12501 points.

In this paper, we shall focus on results pertaining to the analysis of a restricted number of profiles (results on other profiles are comparable):

- 3 tactiles profiles, with code names *BOU* (friction coefficient 0.48), *LRA* (friction coefficient 0.63) and *GRA* (friction coefficient 0.775).
- 2 locked laser profiles, *BOU* again and another profile denoted *CLE* (friction coefficient 0.55).
- 3 non-locked laser profiles, *BOU*, *CLE* and *QB* (friction coefficient 0.65).

As is apparent, the profiles in each acquisition procedure have clearly different friction properties. The friction for tactiles samples ranges from 0.48 to 0.775, while for laser, it ranges between 0.48 and 0.65. Besides checking the fractal behaviour of the profiles, our main aim is to investigate whether fractal parameters are able to

discriminate between profiles with different friction coefficients. As an illustration, we show on figure 1 typical tactile profiles in the classes *BOU*, *LRA* and *GRA*. Figure 2 displays typical *BOU* and *CLE* profiles in the locked laser mode

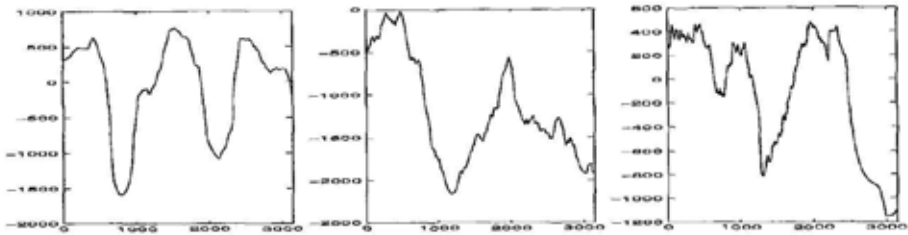


Figure 1. Typical *BOU*, *LRA* and *GRA* tactile profiles.

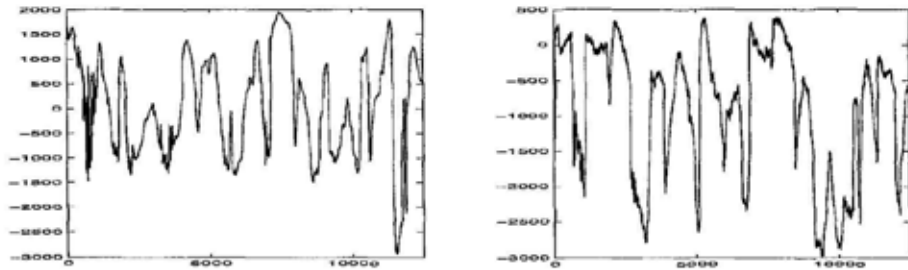


Figure 2. Typical *BOU* and *CLE* laser profiles.

### 3 Fractal analysis

We present briefly in this section the tools that we shall use to perform a fractal analysis of the profiles.

#### 3.1 Continuous wavelet transform (CWT)

Recall that a wavelet is simply a function  $\psi \in L^2(\mathbb{R})$  such that  $\int_{\mathbb{R}} \psi(t) dt = 0$ . Usually, one requires in addition that  $\psi$  be well localized in time and frequency, and has enough vanishing moments (i.e.  $\int_{\mathbb{R}} x^i \psi(x) dx = 0$  for  $i = 1 \dots n$ ).

**Définition 3.1.1.** The continuous wavelet transform<sup>5</sup> of a function  $f \in L^2(\mathbb{R})$  is:

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

As is well known, many fractal properties are related with the evolution of the wavelet coefficients  $CWT(a, b)$  across scale, i.e. with respect to  $a$ .



### 3.2 Correlation exponent

A classical fractal parameter we shall deal with is the correlation exponent. This exponent measures the speed of decay of the autocorrelation of a signal<sup>6</sup>. More precisely, assuming that  $X(t)$  is stationary, denote  $C(l) = E(X(t)X(t+l))$ .

The signal  $X$  has a well defined correlation exponent if  $C(l) \sim l^{-\beta}$  with  $\beta > 0$  holds across a range of values of  $l$ . A particular case is of special interest: The signal  $X$  is said to be long range dependant (LRD) if  $C(l) \sim l^{-\beta}$  when  $l$  tends to infinity, with  $\beta > 1$ . The definition of LRD corresponds to the situation where the series  $\sum_{l \in \mathbb{Z}} C(l)$  diverges.

### 3.3 Regularization dimension (DimR)

Fractional dimensions are one of the best known parts of fractal analysis. In this work, we shall deal with the so-called regularization dimension<sup>7</sup>. A heuristic explanation of DimR is the following. Start with a compactly supported signal  $X$ . For a given positive  $s$ , consider the convolution of  $X$  with a Gaussian kernel of variance  $s$ . Let us denote by  $X_s$  this regularized signal. Assume that  $X$  is so irregular that it has infinite length. Since  $X_s$  is  $C^\infty$  for any positive  $s$ , it has finite length  $L_s$ . Furthermore,  $X_s$  tends to  $X$  when  $s$  tends to zero. The regularization dimension measures the speed of convergence of  $L_s$  to infinity when  $s$  tends to 0 (see figure 3 for an illustration on a road profile).

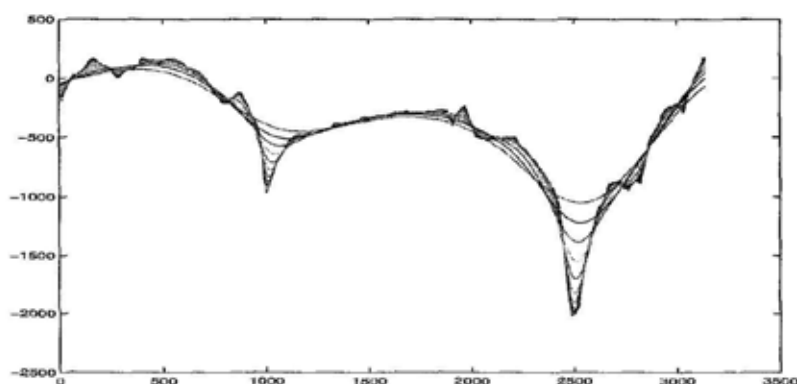


Figure 3. Computation of a regularization dimension on a profile through successive convolutions.

Let us now give a formal definition. Let  $\Gamma$  be the graph of a bounded and compactly supported function  $f : K \subset \mathbb{R} \rightarrow \mathbb{R}$ . Let  $\chi(t)$  be a kernel in the Schwartz class, and set, for  $a > 0$ ,  $\chi_a(t) = \frac{1}{a}\chi(\frac{t}{a})$ . Let  $f_a = f * \chi_a$ . This function is infinitely smooth, and the length of its graph  $\Gamma_a$  on  $K$  is given by

$$L_a = \int_K \sqrt{1 + (f'_a(t))^2} dt$$

**Définition 3.3.1.** *The regularization dimension of (the graph of)  $f$  is:  $\text{Dim}R(\Gamma) = 1 + \limsup_{a \rightarrow 0} \frac{\log(L_a)}{-\log(a)}$*

The regularization dimension coincides in many cases with the better known box dimension: This is in particular the case for all "classical" fractal signals such as Weierstrass functions, fractional Brownian motions, .... One may prove that the relation  $\text{Dim}R \leq \text{Dim}B$ , where  $\text{Dim}B$  is the box dimension, holds for any continuous function. This indicates that  $\text{Dim}R$  is "finer" than the box dimension. The main advantage of the regularization dimension is that it leads to much more accurate estimations on sampled data than the box dimension. This is mainly because the number of smoothed versions is not restricted. Another reason is that  $\text{Dim}R$  is less sensitive to noise than  $\text{Dim}B$ . The interested reader may consult<sup>7</sup> for more details.

From a practical point of view, a signal will be considered fractal if a plot of  $\log(L_a)$  versus  $\log(a)$  is linear in a certain range of values of  $a$ .

### 3.4 Pointwise Hölder exponent

In contrast with the correlation exponent and  $\text{Dim}R$ , which are global quantities, the pointwise Hölder exponent  $\alpha$  measures a local behaviour<sup>8</sup>. Its definition reads:

**Définition 3.4.1.** *Let  $x_0 \in \mathbb{R}$ , and  $s$  be a real number with  $s > -1$ . A function  $f: \mathbb{R} \rightarrow \mathbb{R}$  belongs to  $C_{x_0}^s$  if and only if there exist a constant  $C$  and a polynomial  $P$  of degree at most  $[s]$  such that*

$$|f(x) - P(x - x_0)| \leq C|x - x_0|^s. \quad (1)$$

The pointwise Hölder exponent of  $f$  at  $x_0$ , denoted by  $\alpha_f(x_0)$  or simply  $\alpha$ , is defined to be  $\sup\{s : f \in C_{x_0}^s\}$ .

When  $0 < \alpha < 1$ , it is given by the simple formula:

$$\alpha = \liminf_{h \rightarrow 0} \frac{\log |f(x_0 + h) - f(x_0)|}{\log |h|}$$

Since  $\alpha$  is defined at each point, one may associate to  $f$  its Hölder function:

**Définition 3.4.2.** *Let  $f$  be a bounded function. The Hölder function of  $f$  is the function which associates, to each  $x$ ,  $\alpha_f(x)$ .*

While the Hölder exponents and the Hölder function cannot tell whether a signal is "fractal", they provide a rich description of the local singularity structure of a signal. A small  $\alpha_f(x)$  means that  $f$  is irregular at  $x$ , and vice versa. For instance, if  $f$  is  $C^1$  at  $x$ , then  $\alpha_f(x) \geq 1$ ; If  $\alpha_f(x) < 0$ , then  $f$  is discontinuous at  $x$ .

## 4 Results

We have computed the parameters described in the previous section on our road profiles. All the programs we have used are available in the software toolbox called *FracLab*. *FracLab* may be downloaded at [www.irccyn.ec-nantes.fr/hebergement/FracLab/](http://www.irccyn.ec-nantes.fr/hebergement/FracLab/) and <http://fractales.inria.fr>.

In the next subsection, we verify that the profiles display a fractal behaviour. Then, we use this property to characterize the signals.

#### 4.1 Fractal behaviour

##### Energy and CWT

One possible way to check for a fractal behaviour is to investigate the evolution of the energy in the signal with respect to scale. More precisely, let  $E_a^2 = \int [CWT(a, b)]^2 db$  denote the energy at scale  $a$ . A relation of the type  $E_a^2 \sim a^\gamma$  for some  $\gamma$  and  $a$  in a given range indicates that the energy decays as a power law in scale.

Figure 4 shows that such a relation is approximatively verified for most tactiles profiles across a large range of scales. Results on the other types of profiles are similar.

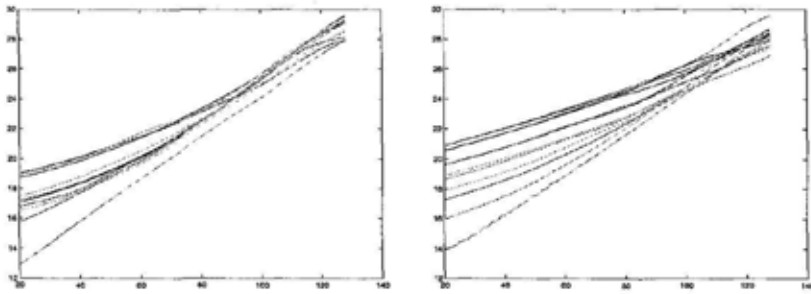


Figure 4. log-energy with respect to scale for the tactiles profiles *BOU* and *LRA*.

##### Correlation exponent

The results for the correlation exponent confirm the ones above. Figure 5 shows that, for tactiles profiles, the logarithm of the lag  $l$  correlation  $C(l)$  behaves linearly as a function of  $\log(l)$  on almost all the range of possible values of  $l$ . Again, the same type of graphs are obtained with other profiles.

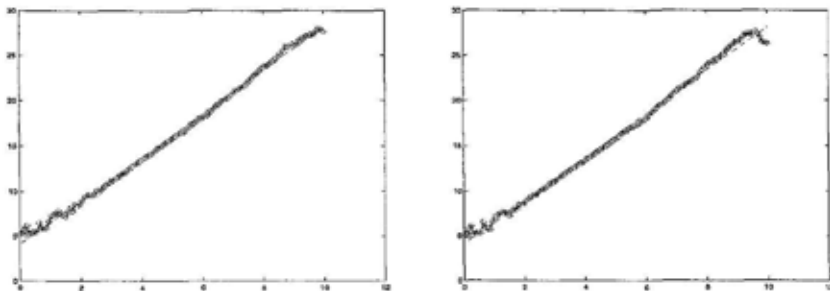


Figure 5. log-correlation as a function of the logarithm of the lag for the tactile profiles *LRA* (left) and *GRA* (right).

From a numerical point of view, the values of the exponents measured on various profiles range between 0.8 and 1.4. Thus, although a clear fractal behaviour is

verified by all profiles, only some of them display LRD.

### Regularization dimension

Figure 6 displays a typical behaviour of DimR on profiles. As one can see, the graph  $\log(L_a)$  versus  $\log(a)$  is not linear. There are however two linear regimes, one corresponding to a low regularization (i.e. high frequencies), and the other one valid for large smoothing, i.e. low frequencies. From the DimR point of view, it thus seems that our road profiles have two well-defined dimensions, indicating that the micro- and macro-textures behave in a different way.

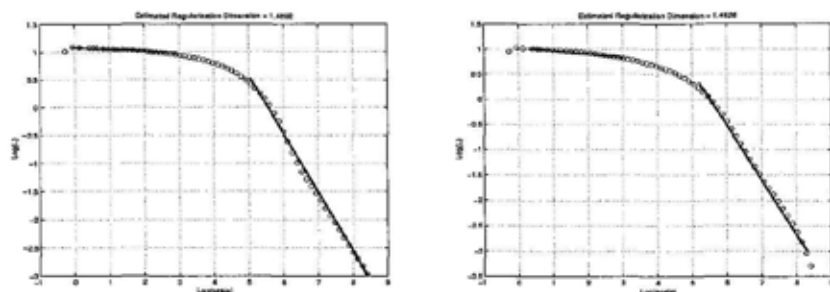


Figure 6. Evolution of the logarithm of the length of smoothed version of the tactile profiles *BOU* (left) and *LRA* (right) as a function of the logarithm of the smoothing parameter.

We note that the regularization dimensions range between 1.1 and 1.6 on our profiles. This indicates that the profiles vary between almost smooth and somewhat irregular.

### 4.2 Profile characterization

Although the profiles all clearly manifest a fractal behaviour, we have found that neither the correlation exponent nor the regularization dimension were able to characterize a given class of profiles. As a consequence, these parameters may not be used to explain the differences in friction of the various profiles. This is seen in a qualitative way on figure 7. The correlation exponents are represented for all tactile profile in the class *BOU*. Though all profiles show an excellent linear behaviour, the slopes of the 15 different samples vary a lot. Thus there is no single exponent that may be meaningfully attributed to a given class. Moreover, a quantitative analysis shows that the ranges of exponents for the different classes overlap a lot. It is thus not possible to separate the classes based on the information brought by the correlation exponent. The same comments apply to the regularization dimension.

This leads us to the following conclusion: If the fractality of the profiles is of any relevance for friction, this should be sought in local features rather than in global ones. Such a claim is supported by the fact that friction is mainly a local phenomenon. As a consequence, global measures of irregularity such as DimR or correlation exponents may be largely unrelated to the friction coefficient. In contrast, local regularity measures such as Hölder exponents should be strongly correlated with friction. We now proceed to investigate such correlations.

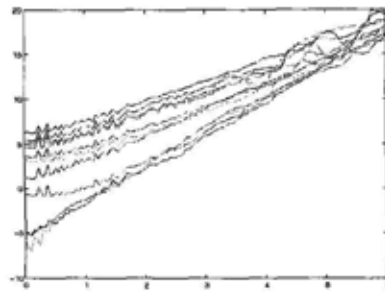


Figure 7. Correlation analysis of the 15 *BOU* tactile profiles.

### Hölder function

As an illustration, the Hölder functions for samples of the tactile profiles *BOU* and *LRA* are displayed on figure 8. Hölder functions for other profiles look similar.

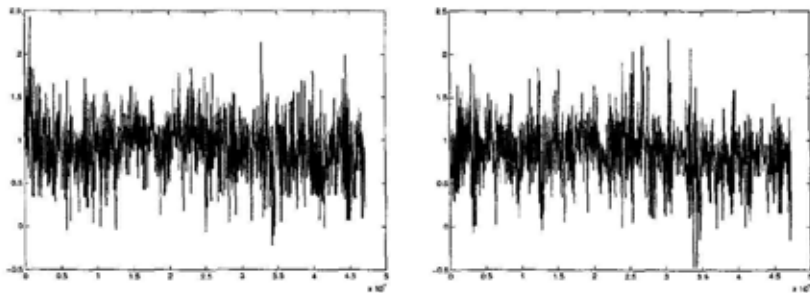


Figure 8. Hölder functions for the tactile profiles *BOU* and *LRA*.

The Hölder function yields too rich an information for our purposes. We start by investigating the use of its median for profile characterization. Note that, while the median will subsume information pertaining to the whole signal, it is still a local parameter. It is thus radically different from a global parameter such as *DimR*. The use of the median (or the mean) of the Hölder function is consistent with the fact that the friction results from an average of many local interactions.

### Combination of *DimR* and the Hölder median

We first compute the Hölder functions of all the profiles and all the samples. We then extract their median. Figure 9 shows an attempt to classify the different classes based on this median plus the regularization dimension. While this procedure works well for the two laser profiles in locked mode, it fails to separate the three classes of laser profiles in unlocked mode.

We now discuss another technique that makes a fuller use of the information brought by the Hölder function.

### Histograms of Hölder functions

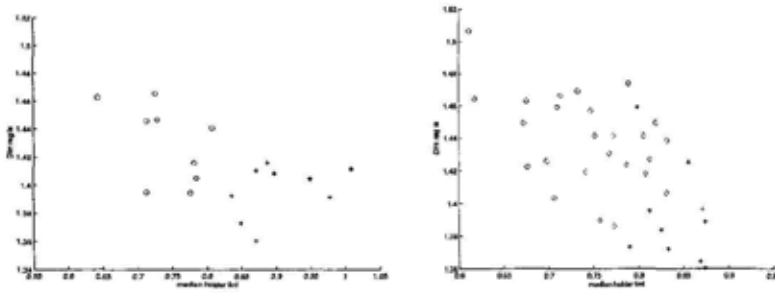


Figure 9. Classification of the profiles classes using the Hölder median (abscissa) and DimR (ordinates). The left plot displays locked mode laser profiles *CLE* (circles) and *BOU* (stars). The right plot shows unlocked mode laser profiles *CLE* (circles), *QB* (diamonds) and *BOU* (stars).

Instead of restricting to the median, we now study the histograms of the Hölder functions. More precisely, for each profile  $P$ , we compute ten histograms  $H_P(\alpha)$  of the Hölder functions evaluated for ten "test" samples randomly chosen among the fifteen samples in profile  $P$ . We do this for the two locked mode laser profiles *BOU* and *CLE*, and for the three unlocked mode laser profiles *BOU*, *CLE* and *QB*. See figure 10.

The second step is to model these histograms as Gaussian processes. In other words, for each given profile and for each value of  $\alpha$ , we compute the mean  $M_P(H_P(\alpha))$  and variance  $V_P(H_P(\alpha))$  of  $H_P(\alpha)$  evaluated over the ten test samples in this profile. Let  $G_P(H_P(\alpha)) = \mathcal{N}(M_P(H_P(\alpha)), V_P(H_P(\alpha)))$  denote the Gaussian distribution obtained for the histogram of the profile  $P$  at value of the exponent equal to  $\alpha$ . To check whether a new, unknown, sample belongs to profile  $P$ , one first computes the histogram  $h(\alpha)$  of its Hölder function. If the sample belong to  $P$ , we expect that  $G_P(h(\alpha))$  is "large". A quantity that measures how the sample is "close" to profile  $P$  is thus:

$$T_P(h) = \int_{\mathbb{R}} G_P(h(\alpha)) d\alpha$$

Any unknown sample is then attributed to the profile  $P$  which maximizes  $T_P(h)$ .

We found that this method was able to classify with 100% success the five remaining samples in all classes. Our conclusions are thus as follows:

- Road profiles indeed display fractal behaviours in terms of both correlation exponent and regularization dimension over a large range of scales.
- Global fractal measures as are the correlation exponent and regularization dimension do not allow to characterize profiles.
- The local regularity information brought by the Hölder exponent allows to classify the profiles through a simple statistical procedure.

Future work will focus on relating the structure of the Hölder function with the friction coefficient of the profiles.

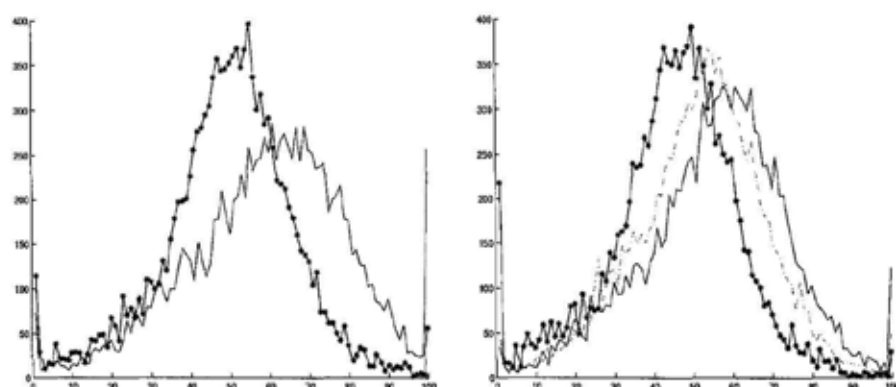


Figure 10. Left: Mean of the histograms of the Hölder functions for locked mode laser profiles *BOU* (solid line), *CLE* (stars). Right: Same, but for unlocked mode laser profiles *BOU* (solid line), *CLE* (stars) and *QB* (dotted).

## References

1. Z. RADO, A Study of Road Texture and its Relationship to Friction, Phd Thesis, The Pennsylvania State University, **1994**.
2. G. HEINRICH, Rubber friction on self-affine road tracks, *Rubber Chemistry and Technology* 70, **1997**.
3. M.T. DO, H. ZAHOUANI, Frottement pneumatique/chaussée, influence de la microtexture des surfaces de chaussée, *JIFT*, **2001**.
4. M. KLÜPPEL, Rubber friction on self-affine road tracks, *Rubber Chemistry and Technology* 73, 578-606, **2000**.
5. YY. MEYER, *Ondelettes et Opérateurs*, volume 1. Hermann, Paris, **1990**.
6. J. BERAN, *Statistics for Long-Memory Processes*, Chapman and Hall, New York, **1994**.
7. F. ROUEFF, J. LEVY VÉHEL, A regularization approach to fractionnal dimension estimation, *Fractal conference*, **1998**.
8. K. DAOUDI, J. LÉVY VÉHEL, Y. MEYER, Construction of Continuous Functions with Prescribed Local Regularity, *Constructive Approximation*. Vol. 14. Num. 3, pp.349-386, **1998**.
9. M. GUGLIELMI, J. LEVY-VEHEL, Analysis and simulation of road profile by means of fractal model, *Conference on Advances in Vehicle Control and Safety (AVCS 98)* Amiens, **1998**.

# FRactal Distributions of Temperature, Salinity and Fluorescence in Spring 2001-2002 in South San Francisco Bay

KAREN FISHER

*PO Box 1663; MS B296; Los Alamos National Laboratory, Los Alamos NM 87545, USA*

*Email: kefisher@lanl.gov*

WIM KIMMERER

*3152 Paradise Drive; Romberg Tiburon Center, San Francisco State University, Tiburon CA 94920, USA*

*Email: kimmerer@sfsu.edu*

In this paper, we demonstrate a wavelet-based analysis of the changing fractal character of temperature, salinity, and fluorescence distributions in South San Francisco Bay. In particular, we are interested in comparing the fractal character of the physical and biological elements of the system. This analysis indicates that the system exhibits two distinct states; one in which physical (salinity and temperature) and biological (fluorescence) variables show similar fractal character across scales of 120 meters to 2 kilometers, and one in which fluorescence shows substantially different character than the others. These regimes occur in spring seasons with an episodic fluorescence structure (2002), and a more consistent seasonal structure (2001) respectively. During spring of 2001 fluorescence has somewhat higher persistence than the physical variables. During spring of 2002, the physical and biological variables have similar persistence. Differences in the response of fluorescence to physical variables may reflect differences in the pattern of stratification and bloom development during these years. Analysis of the fractal character of these signals offers a promising approach to assessing and adequately modeling the patchiness of biological distributions at scales of meters to kilometers in rapidly fluctuating dynamic systems. Realistic patchiness can then be used in models of the system, leading to better characterization of biological and physical coupling across a wide range of scales.

## 1 Introduction

Many oceanographic research programs measure chlorophyll fluorescence while ships are underway to provide a rapid index of phytoplankton distribution. High-frequency spatial or temporal variability in fluorescence has been used to infer underlying patterns and causes of variation in phytoplankton biomass [20]. In shallow estuaries, phytoplankton biomass is controlled by three principal factors: the availability of nutrients for growth [18], penetration of light into the water [7], and grazing by benthic organisms, particularly bivalves [1, 2, 19]. Mixing of water masses with variable chlorophyll concentration can result in complex patterns of variability in fluorescence [10].

In South San Francisco Bay, nutrient concentrations are generally high, and phytoplankton blooms can develop when the water column stratifies [3]. South Bay is usually vertically well-mixed, except during times of low tidal energy when stratification can develop and persist over several days [3]. Stratification commonly forms and dissipates at the tidal time scale [17], but this is too rapid for phytoplankton to respond [12]. Persistent stratification arises during times of low tidal energy (neap tides), and reaches peak levels when Central Bay is relatively fresh because of extreme outflow



events [3, 6]. Persistent stratification during the spring results in a strong phytoplankton bloom [3, 4, 6].

Stratification affects bloom formation by trapping phytoplankton in the surface layer, enhancing light availability and isolating the developing bloom from benthic grazers, which can otherwise be a key sink for phytoplankton biomass [2]. However, bloom formation in the South Bay is spatially as well as temporally variable. Because of the strongly variable spatial gradients, variation in residual currents due to wind, local runoff or other subtidal effects can be important in distributing and dispersing the bloom [10, 21]. Consumer organisms including bacteria, zooplankton, and clams respond to these blooms with increased growth rate or biomass [5]. Furthermore, the actual process of bloom formation is complicated by physical transport, and particularly by interactions between shoals and channels [13, 14].

Freshwater flow has two modes of influence on South Bay. First, local streams and waste-water treatment plants introduce freshwater at the southern end of South Bay, reducing salinity. Nutrient inputs in South San Francisco Bay primarily come from wastewater treatment plants [9]. Second, high freshwater flow into the northern estuary reduces salinity in the Central Bay, setting up inverse estuarine circulation in South Bay, with residual circulation to the south at the surface and north at the bottom [16]. This promotes stratification and decreases water residence time in the South Bay [23].

Here we examine data gathered along the track line of the USGS research ship R/V *Polaris* for 26 cruises conducted between 2000 and 2002. Our goal was to use available information about spatial structure to better understand spring bloom dynamics in South San Francisco Bay. Since fluorescence can be measured relatively rapidly, its spatial distribution provides a reasonably comprehensive index for the spatial distribution of phytoplankton [11]. Due to the multiplicity of interacting forces that determine the distribution of phytoplankton, and the range of scales over which they operate, the information gained from the fluorescence signal along the ship track is invaluable for understanding the patch structure of phytoplankton at scales smaller than the spacing between fixed stations. In particular, we are interested in the times and spatial scales at which the distributions of salinity, temperature, and fluorescence are similar, and when they are different.

## 2 Data and Methods

Fluorescence, salinity, and temperature data were obtained from a transect along the shipping channel in South San Francisco Bay (Figure 1) from 26 cruises between August 2000 and May 2002. The along-track records were analyzed to determine the spatial scale of patchiness of chlorophyll fluorescence, and the relationship of this biological variable to the physical variables of salinity and temperature. Two sets of cruise records are shown in Figure 1, taken on April 16<sup>th</sup> of 2001 and 2002. Because of the importance of the bathymetry along the shipping channel to the location of various features in the surface fluorescence concentrations, the bathymetric profile is shown for comparison. All three variables have a considerable degree of variability, both along the records spatially, and between the records for neighboring cruises (Figure 2). The patterns for both temperature and salinity show a smoother, more seasonal transition between saltier colder water and fresher warmer water in spring 2001 than in spring 2002.

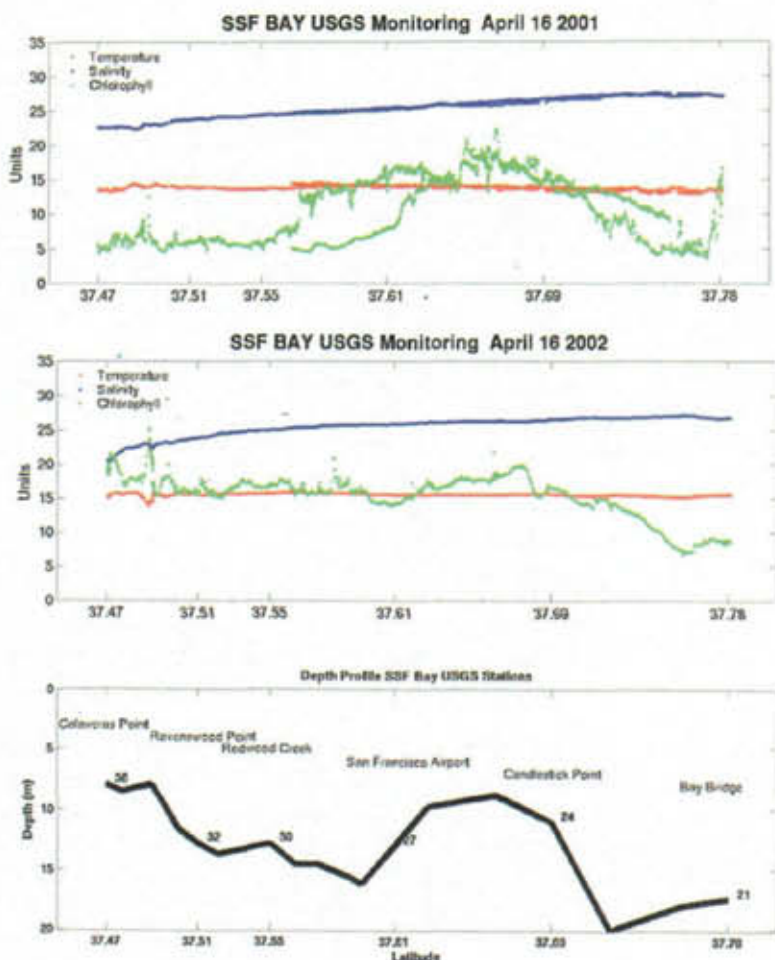


Figure 1.) Salinity (blue), temperature (red) and fluorescence (green) records for two spring cruises in South San Francisco Bay in April 2001 (top) and April 2002 (middle). Depth profile between USGS numbered stations along the cruise track (bottom).

Because of the large influence of processes occurring at tidal timescales, it is difficult to estimate the duration of structural changes in the temperature and salinity. Tidal variation can easily cause substantial changes in density structure. During these cruises, every density profile taken on April 16<sup>th</sup> 2001 is denser than its corresponding station profile taken in 2002, and with the exception of the southern most station, every 2001 profile has a larger gradient in density from the surface to the bottom. At the times sampled, all seventeen station profiles taken in South Bay on April 16<sup>th</sup> 2001 exceed a commonly used criterion for significant density change (0.125) by a depth of 8 meters, while on April 16<sup>th</sup> 2002 only eight profiles exceed the criterion, six of them at 8 meters or deeper. These differences in density states may influence the patterns seen in these snapshots somewhat, but it is expected that whether these states exist for hours or days has an even larger influence on the scales of patchiness observed.

## 2.1 *Wavelet-based variance spectra characterizing along-track variability*

Spectral analysis has a long history as an approach to analyzing the spatial structure of plankton, dating back to over three decades (e.g. [20]). The relative slopes  $\beta$  of the spectra give the relationship of the biological (fluorescence) and physical (salinity and temperature) variables that yield insight into the shifting dynamics governing the standing stock of chlorophyll [8]. In other words, the pattern gives clues to the process. The simplified rules of spectral analysis are as follows.

An abrupt break in slope indicates a change in the dynamics governing a distribution at a particular spatial scale. A peak in a spectrum can result from a dominant patch size, for example if every fluorescence peak encountered was exactly 500 meters long along the track line, a peak would result at 500 meters in scale in the power spectrum. In general, a slope near 2 is consistent with isotropic (non-directional) physical controls. A slope near 3 is consistent with 2-D (e.g., generated by tidal friction) physical controls. A slope near zero is almost random, consistent with a well-mixed system with no remaining evidence of stirring or eddies. Slopes near 1 tend to result in systems where zooplankton or other grazers are present and active, or where phytoplankton growth or losses are occurring. Phytoplankton spectra can have a stair-step shape, with steep and flat sections alternating [e.g. 8]. Steeper sections tend to have slopes consistent with physical control. An exception to this is that in a well-mixed system, where physical variables have flat spectra, a steep biological spectrum may result from biological controls which result in localized sources or sinks (e.g., a rapidly growing patch of phytoplankton, or a concentration of efficient grazers).

Malamud and Turcotte (1999) review the methods available for analysis of the frequency domain (spectral analysis), and subsequent fractal characterization of natural data. They conclude that because wavelets are localized, and can be directly applied to data from anisotropic fields, wavelet-based spectra are a preferable approach. Wavelet-based spectra can also be obtained from non-stationary fields where, broadly speaking, the mean or the variance changes with the length of the interval considered. Wavelet analysis is therefore a powerful way to analyze the spatial structure of a highly variable system such as South San Francisco Bay. Spatial scales of plankton variability have been noted over many orders of magnitude [20]. Power-law distributions have been shown in fluorescence distributions [20], forming a straight line on a log-log plot of variance vs. scale that shows these distributions to be self-affine, or fractal. Fractal character determined from the wavelet-based spectra reflects both the heterogeneity and the variability of an observed subset of each system [8].

## 2.2 *Wavelet Transform and Resulting Power Spectra*

The wavelet transform is a filter  $g$  passed over a data series  $f(x)$ . A wavelet transform used for data analysis is often implemented in a form becoming known as "semi-discrete", because the transform is calculated at an arbitrary number of discrete scales using an approach similar to that used in a continuous transform. The generalized form of a semi-discrete one-dimensional wavelet transform is:

$$W(x, a_n) = \frac{1}{\sqrt{a_n}} \int_{-\infty}^{\infty} g\left(\frac{x' - x}{a_n}\right) f(x') dx' \quad (1)$$

where  $f$  is the function of the spatial position  $x$ ,  $a_n$  is the scale being analyzed, and  $g$  is the filter [15]. The number and distribution of dilations  $n$  are chosen to adequately resolve the spatial scales of features in the dataset, often chosen to be powers of 2 ( $n=1, 2, 4, \dots$ ) or integers ( $n=1, 2, 3, \dots$ ) depending on the resolution required. A variety of filters have been developed and tested, with different advantages in discriminating features of the data [e.g. 25]. The only requirement for designing a wavelet filter is that it must be continuous and integrate to zero:

$$\int_{-\infty}^{\infty} g(x') dx' = 0 \quad (2)$$

The Mexican Hat filter has been used to determine features of naturally occurring distributions and the general shape of the Mexican Hat filter resembles frontal features encountered in the along-track data obtained during research cruises. The Mexican Hat filter has the form:

$$g(x') = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} (1 - x'^2) e^{-\frac{x'^2}{2}} \quad (3)$$

Substituting equation 3 into 1 gives the following form [15]:

$$W(x, a_n) = \left(\frac{1}{2a_n\pi}\right)^{\frac{1}{2}} \int_{-\infty}^{\infty} \left[1 - \left(\frac{x' - x}{a_n}\right)^2\right] e^{-(x' - x)^2 / 2a_n^2} f(x') dx \quad (4)$$

The amplitude of the wavelet transform  $W(x, a_n)$  at each scale  $a_n$  reflects the relation between adjacent areas at that scale; it measures the intensity of each contributing frequency. For white noise with values equally uncorrelated at all scales, the wavelet magnitudes will not change with scale. Conversely, in eddies, the energy associated with larger spatial scales will be higher as the wavelet encounters persistent peaks and valleys on a scale similar to the wavelet scale.

Following the analysis of Malamud and Turcotte (1999), self-affine series can be characterized succinctly because such series have a power-law dependence of the power spectral density function on frequency  $f$ :

$$S(f) \propto f^{-\beta} \quad (5)$$

The exponent  $\beta$  is the negative slope of the power spectrum of the series. The relation in equation 5 defines a self-affine fractal in the same way a self-similar fractal is defined [15]. The variance  $V_n$  of the wavelet transform similarly has a power-law dependence on scale  $a_n$  (inverse frequency):

$$V_n \propto a_n^\beta \quad (6)$$

Calculating the variance  $V_n$  of the wavelet transform at each scale  $a_n$ , and plotting it as a function of scale examined (not frequency) in log-log space produces a power-law exponent  $\beta$ . The exponent  $\beta$  produced in wavelet variance analysis has the same magnitude, but opposite sign, from Fourier power-spectral analysis performed after detrending and windowing.

### 3 Results

The wavelet transforms for along-track distributions of fluorescence on April 16, 2001 and April 16, 2002 are shown in Figure 2. This figure is illustrated with amplitude squared for direct comparison of each location along the track line to the wavelet spectra that are integrated over the entire region and shown in Figure 3. Flat regions in the far (2.2 km scale) corners of the plots in Figure 2 at either end of each transect show the area masked by the effects of edges. Note especially the following features of these two wavelet magnitude plots. First, the general impression of the distributions differs between the two years. The spatial location of the peak in fluorescence concentration changes between the two years, as does the amount of spikiness in the wavelet transform. Moreover, the April 16, 2001 plot shows some locations where the maximum amplitude of the wavelet transform is at smaller scales, a departure from the general expectation that power will "cascade" from large scales to small scales. This indicates that variance is being produced at these smaller scales, i.e., in local areas, and is evident in the spectra for temperature and salinity for this cruise (Figure 3).

A local peak in the transform value will occur at the scale where the peak of the local variance is the size of a local "patch" or front. Only scales for patches that occur at some characteristic size throughout a region show up as peaks in the regional power spectrum constructed from the wavelet variance spectra along the whole track line (Figure 3). The April 16, 2002 plots for both Figure 2 and Figure 3 show that most of the peaks are located at the larger scales of analysis. These results suggest that the system was in different states during these two cruises. Spectra are also shown for the cruises immediately following these two target cruises in Figure 3. Of particular note, although the spectral shapes and relative magnitudes of the April 22<sup>nd</sup> 2002 cruise are still similar to those of the preceding cruise, the slope  $\beta$  of the smaller scale fluorescence has flattened dramatically. The overall tendency of the spectra throughout the spring of 2002 to be coupled in shape for all three spectra is also illustrated, relative to the spectra in the spring of 2001 having distinct shapes of the fluorescence spectra.

Spectra for April 16<sup>th</sup> and 26<sup>th</sup>, 2001 are similar in magnitude but decoupled in shape. Although total variance in fluorescence, salinity, and temperature are very close to the same throughout this spring, the differences in spectral shape are ubiquitous. The "inverse cascade" in salinity and temperature, whereby the spectrum drops off at larger scales in the earlier cruise, indicates that there is patch scale of about 700 meters dominating the physical variance in this spatial range. Fluorescence shows a drop in variance between 1 and 2 km in the latter cruise not seen in the physical variables at that time. For April 16<sup>th</sup> and 22<sup>nd</sup> 2002, the reverse is true: the spectra are decoupled in total variance, with several orders of magnitude less variance in salinity and temperature than

in fluorescence, but all three spectra have parallel trends (similar shapes), although fluorescence has a somewhat flatter slope particularly during the later cruise.

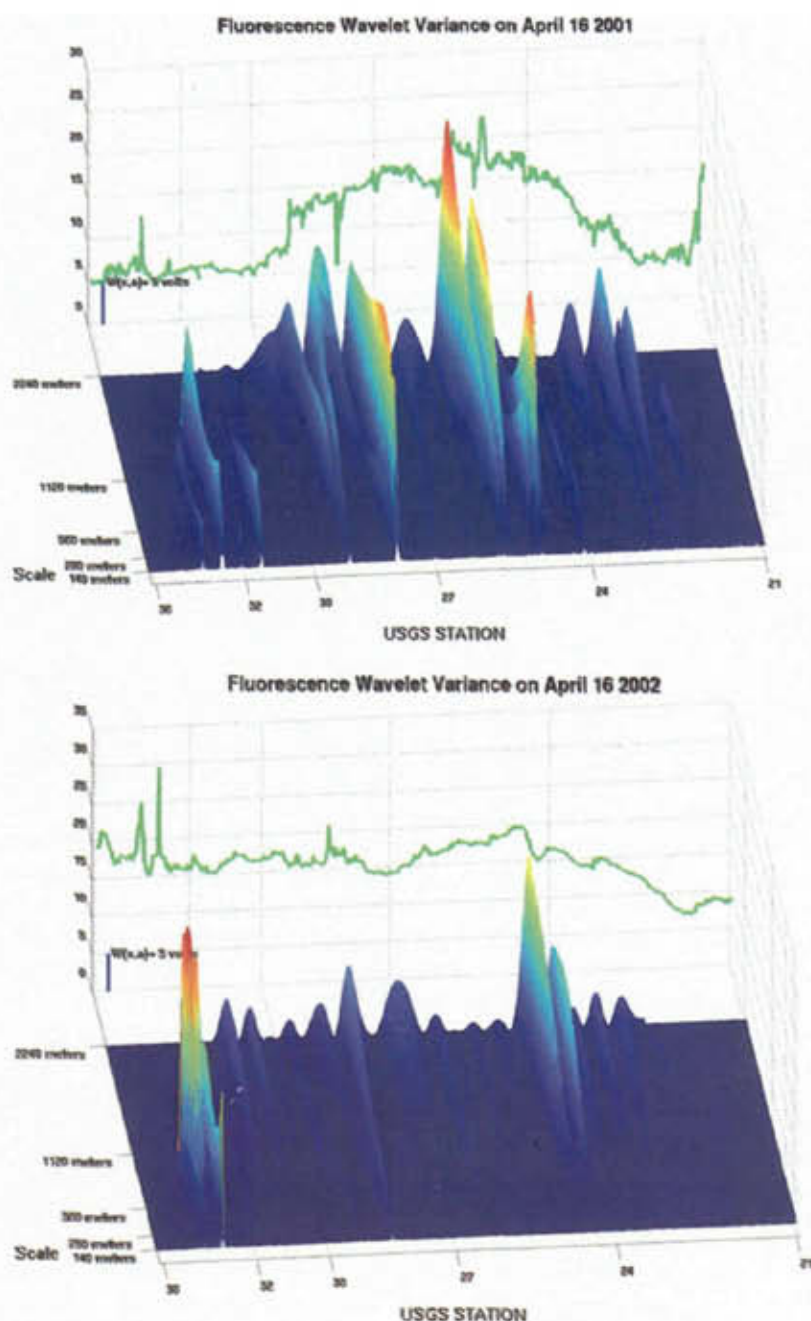


Figure 2.) Absolute values of the magnitude of the Mexican Hat wavelet transform, shown along the track line of South San Francisco Bay for April 16<sup>th</sup> 2001 and 2002, respectively. USGS stations run from station 36 at the southern end of South San Francisco Bay, to station 21 near the boundary with Central San Francisco Bay.



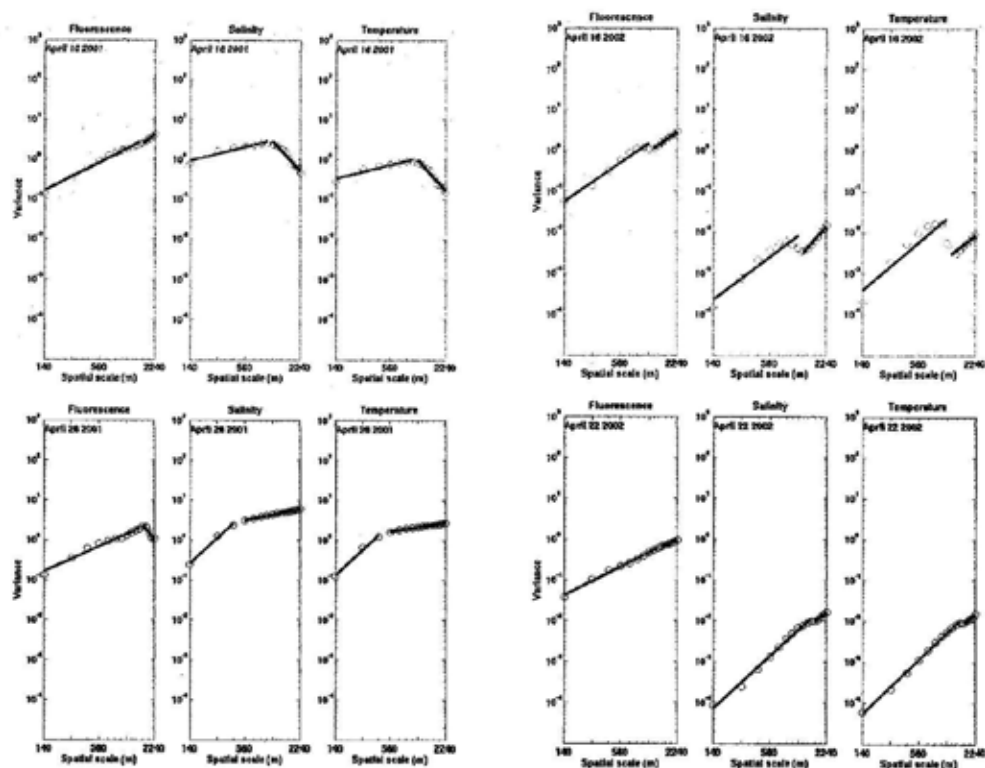


Figure 3.) Wavelet-based power spectra for fluorescence (left), salinity (middle) and temperature (right) records on April 16<sup>th</sup>, 2001 and 2002 (top) and their successors on April 26<sup>th</sup> 2001 and April 22<sup>nd</sup> 2002 (bottom).

The relationships among slopes of temperature, salinity, and fluorescence fall into two statistically distinguishable categories of biologically-physically coupled and uncoupled systems. The slopes of all spectra determined for fluorescence, salinity, and temperature along track lines of the USGS survey were compared for each of the 26 cruises. These data are unevenly spaced in time (more intensive sampling in spring), but evenly spaced along transects (at 44 meter intervals).

Throughout the spring of 2001, the slopes of fluorescence were decoupled, tending to move differently from the slopes of salinity and temperature and to break at different scales. In contrast slopes of all three variables were quite coupled in 2002 (Figure 4). Each date label on the x-axis of Figure 4 is color-coded to indicate whether the fluorescence is coupled in shape (red), coupled in both shape and total power (black), or decoupled (green) relative to the physical distributions. The observed density structures' states are extremely different in the two Aprils, and although the duration of the states cannot be directly assessed from these data, the cruises before and after April 16<sup>th</sup> 2001 exhibit similar structure, while those closest to April 16<sup>th</sup> 2002 exhibit less well mixed profiles than this cruise.

Benthic clams are expected to exert the highest grazing pressure during periods of mixing, while stratification isolates the surface water in which fluorescence levels are

measured from this pressure and allow phytoplankton growth. Consistent with this idea, the fluorescence levels in the shallow southern reaches of South Bay rise following the April 16<sup>th</sup> 2001 cruise, and fall following the April 16<sup>th</sup> 2002 cruise (Figure 4). In addition, the wavelet-based fluorescence spectra for the April 22nd cruise in 2002 indicates that the spectral shape has flattened considerably relative to the slopes for salinity and temperature, a result predicted by Powell and Okubo (1994) for a population under grazing pressure in a system driven by 2-D turbulence [22]. Clearly, both the temporal and spatial information in this along-track dataset can yield insight into bloom dynamics not obtained from the station data taken at the six stations in South San Francisco Bay indicated in Figure 1.

#### 4 Discussion

Powell et al. (1989) examined kilometer-scale variability of chlorophyll concentration, salinity, and temperature during spring 1987, and concluded that spatial variability of physical variables did not explain that of chlorophyll [21]. Our results suggest a subtle coupling revealed by an analysis that allows for changes in pattern with distance, and that this coupling is present during some periods and absent in others. The difference in degree of coupling between the two years (Figure 4) is remarkable. Both years had hydrographs of low freshwater flow, with a higher, earlier January peak in 2002 and a lower, later March peak in 2001 (data from Department of Water Resources, not shown).

Data are not available to assess the frequency and duration of stratification in either year, as each cruise provides snapshots of stratification that can change significantly at the tidal timescale [17]. Therefore any influence of stratification, with attendant positive effects on bloom formation [4, 6], can only be inferred from the patterns that are observed. Nevertheless, the large difference in coupling of biological and physical patchiness between these two years suggests differences in biological mechanisms, possibly variability in benthic grazing or phytoplankton growth rates (e.g. [24]).

The indication of a critical scale below the 2 kilometer cut-off in our resolution is notable. This critical scale is indicated by the abrupt change in power law behavior between 500 meters and 1 kilometer. The rapidly changing character of spatial structure at these scales suggests that the balance of processes controlling the construction and maintenance of fluorescence, salinity, and temperature distributions is different at scales less than 500 meters, and at scales greater than 1 kilometer. Resolving the dynamics responsible for the patterns observed requires that surveys address these spatial scales and take into account the existence of the transition region when determining station spacing. In addition, process-based studies concentrating on resolving the dynamics throughout the transition scales are needed. The evidence of strong coupling and decoupling between power-spectral slopes of fluorescence and salinity and temperature throughout entire seasons indicates that resolving the processes connecting the biological and physical distributions within the range of the critical scale is necessary for both model and observational strategies in South San Francisco Bay. Short time scales (tidal cycles) and small spatial scales (hundreds of meters) need to be resolved to dynamically characterize this system.



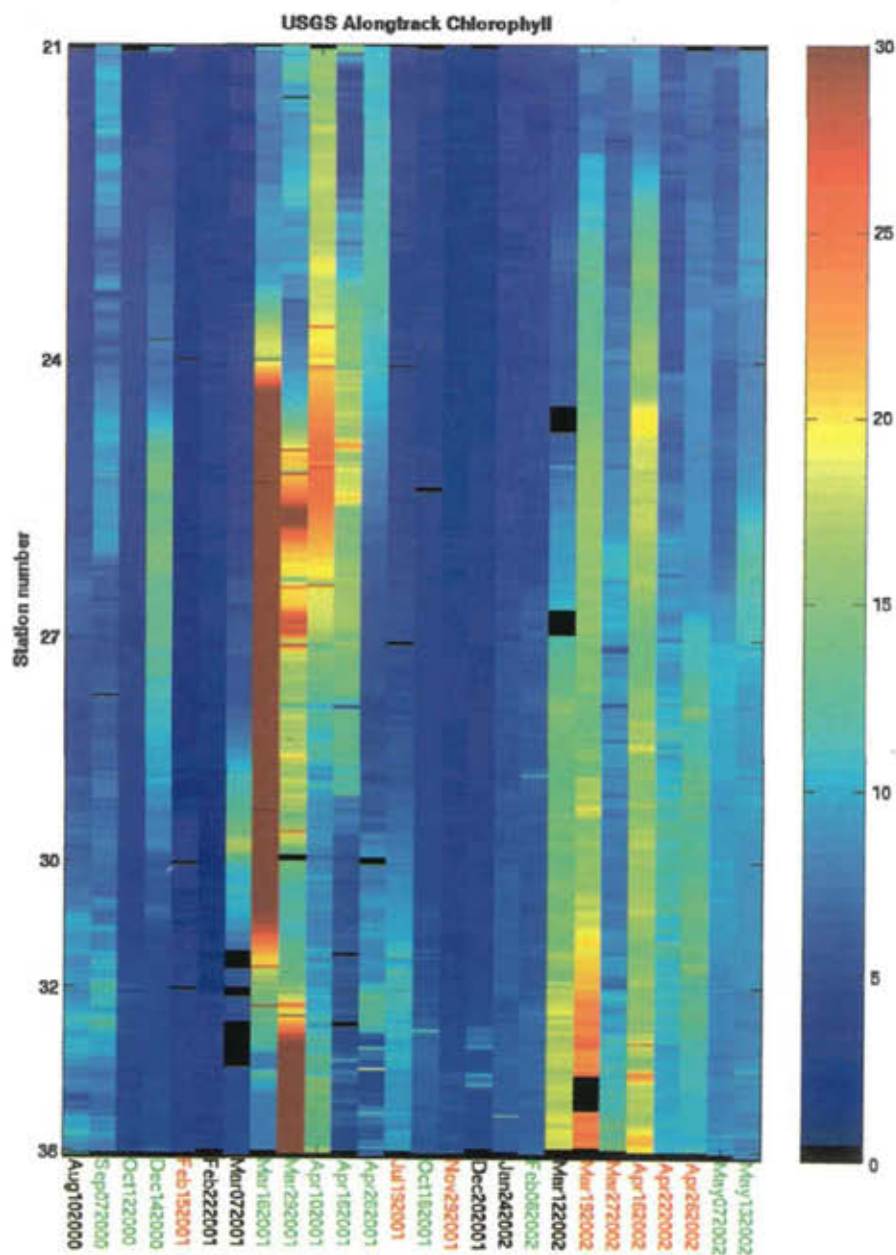


Figure 4.) Fluorescence records analyzed for all 26 cruises, showing seasons along x-axis determined from wavelet analysis to be coupled with temperature and salinity in spectral slope (cruise date in red or black), or decoupled (green)

Several indirect indications of the processes controlling the coupled and decoupled states of Spring 2001 and 2002 in South San Francisco Bay arise from this analysis. Physical structure during a given cruise does not yield sufficient information to connect the patterns seen in fluorescence directly to salinity and temperature, but looking at changes in relative spectral characteristics between cruises reveals possible connections. In particular, the rapid changes in state of April 2002, seen throughout that spring, seem to keep the overall spectral shapes similar. Between April 16<sup>th</sup> and April 22<sup>nd</sup> 2002 changes occur in relative slope  $\beta$  of fluorescence relative to salinity and temperature that are consistent with expectations of population structures in pelagic organisms subject to grazing pressure [22]. The impact of grazing on phytoplankton distributions is a subject with implications for dynamic models of freshwater and marine systems ranging from small ponds to the global ocean. Difficulty quantifying this impact limits the development of predictive models of processes ranging from eutrophication to carbon sequestration. Grazers range in size from microbes to whales and are notoriously challenging to survey across all the relevant size classes during any particular study. Therefore, indirect methods that locate evidence of grazing impact are highly desirable in spite of the clear limitations on their application in the absence of *in situ* measurement. In addition, methods that resolve the spatial scales of variability, such as wavelet analysis, can provide concrete guidelines for effective survey design as an alternative or supplement to statistical approaches (e.g., [11]). Survey designs resolving critical scales where structural changes in power-law distributions occur are highly desirable to maximize the integration of *in situ* measurements with models. In particular, the fractal character of each power-law regime can be used to construct model input fields with realistic variability. In systems where intermittency is substantial, a simple extension of the wavelet-based methods shown above yields multifractal descriptors that can be used to generate realistic portrayal of higher order moments in the model input fields as well.

## 5 Conclusions

In South San Francisco Bay, power law distributions at scales ranging from  $10^1$  to  $10^3$  m show distinct character in different seasonal and hydrographic regimes. The results of the wavelet analysis strongly suggest that scales of variance on spatial scales between 500 meters and 2 kilometers can be dramatically different from that at scales under 500 meters, under certain conditions. When a single power law structure prevails, no change in the structure of variance is observed at scales smaller than 2.24 kilometers, the upper limit of this analysis. At other times, spectral breaks are characteristic of the system at scales smaller than this, indicating that scales of hundreds of meters have significant patchiness that ought to be resolved by more finely gridded sampling strategies. Additionally, it is clear from the along-track data above that rapid variation on both temporal and spatial scales marks the South San Francisco Bay system during some periods. (The movement and reshaping of the fluorescence peak in April 2001 shown in Figure 1 results from changes in the tidal phase between the trip down and the trip back). Processes such as the spring-neap tidal cycle can clearly have impact on this variation. By analyzing the fractal character of these signals, we can assess the states of the system of that result from the confluence of processes underlying the observed patterns in distribution. Fractal descriptors inform model input fields by providing a way to statistically describe the patchiness of the system that is crucial to many of the interactions within the system.

## 6 Acknowledgements

This analysis was performed on preliminary data gathered from the R/V *Polaris* by the Water Quality of San Francisco Bay Project team from USGS (Menlo Park). The analysis was funded by the City of San Jose, as part of an effort to aid in research aimed at understanding the historical dynamics of South San Francisco Bay. Jim Cloern heads the USGS team, which went to extraordinary lengths to make the data available in time for this publication. Tara Schraga invested both time and energy at every level of the data collection and distribution. The authors appreciate that the resources necessary for obtaining the data required for this type of study are extensive, and go far beyond what the authors could accomplish without the dedicated support of the many people involved in the monitoring and assessment of San Francisco Bay, and wish to thank all those we are not able to explicitly mention here for their contributions as well. This document has been assigned LA-UR-03-7509 by Los Alamos National Laboratory.

## References

1. Alpine, A.E., and Cloern, J.E., 1992, Trophic interactions and direct physical effects control phytoplankton biomass and production in an estuary: *Limnology and Oceanography*, v. 37, p. 946-955.
2. Cloern, J. E. 1982. Does the benthos control phytoplankton biomass in South San Francisco Bay (USA)? *Mar. Ecol. Progr. Ser.* 9:191-202.
3. Cloern, J. E. 1984. Temporal dynamics and ecological significance of salinity stratification in an estuary (South San Francisco Bay, USA). *Oceanol. Acta.* 7:137-141.
4. Cloern, J. E. 1991. Annual variations in river flow and primary production in the South San Francisco Bay estuary (USA). Pages 91-96 in M. Elliott and J.P. Ducrottoy, editors. *Estuaries and coasts: spatial and temporal intercomparisons*. Olsen and Olsen, Fredensborg.
5. Cloern, J. E. 1996. Phytoplankton bloom dynamics in coastal ecosystems: a review with some general lessons from sustained investigation of San Francisco Bay, California. *Rev. of Geophysics* 34:127-168.
6. Cloern, J. E., and A. D. Jassby. 1994. Year-to-year fluctuation in the spring phytoplankton bloom in South San Francisco Bay: An example of ecological variability at the land-sea interface. Pages 139-149 in J. H. Steele, T. M. Powell, and S. Levin, editors. *Ecological Time Series*. Chapman Hall, London.
7. Cole, B.E., and Cloern, J.E., 1984. Significance of biomass and light availability to phytoplankton productivity in San Francisco Bay: *Marine Ecology- Progress Series*, 15:15-24.
8. Fisher, K.E., P.H. Wiebe, and B.D. Malamud. 2004. Fractal characterization of local hydrographic and biological scales on Georges Bank. In "Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation". [Eds] Seuront, L. and P. G. Strutton CRC Press. ISBN: 0849313449. Pgs 297-319.
9. Hager, S. W., and L. E. Schemel. 1996. Dissolved inorganic nitrogen, phosphorus and silicon in South San Francisco Bay. I. Major factors affecting distributions. Pages 189-215 in J. T. Hollibaugh, editor. *San Francisco Bay: The Ecosystem*. AAAS, San Francisco.

10. Huzzey, L. M., J. E. Cloern, and T. M. Powell. 1990. Episodic changes in lateral transport and phytoplankton distribution in South San Francisco Bay. *Limnol. Oceanogr.* 35:472-478.
11. Jassby, A.D., Cole, B.E., and Cloern, J.E., 1997. The design of sampling transects for characterizing water quality in estuaries: *Estuarine, Coastal and Shelf Science*, v. 45: 285-302.
12. Lucas, L. V., J. E. Cloern, J. R. Koseff, S. G. Monismith, and J. K. Thompson. 1998. Does the Sverdrup critical depth model explain bloom dynamics in estuaries? *J. Mar. Res.* 56:375-415.
13. Lucas, L.V., J.R. Koseff, J.E. Cloern, S.G. Monismith and J.K. Thompson. 1999. Processes governing phytoplankton blooms in estuaries. Part I. The local production-loss balance. *Marine Ecology Progress Series*, v. 187: 1-15.
14. Lucas, L.V., J.R. Koseff, J.E. Cloern, S.G. Monismith and J.K. Thompson. 1999. Processes governing phytoplankton blooms in estuaries. Part II. The role of transport in global dynamics. *Marine Ecology Progress Series*, v.187: 17-30.
15. Malamud, B.D., and D. L. Turcotte, 1999. Self-affine time series: I. Generation and analysis. *Advances in Geophysics*, 40:1-90.
16. McCulloch, D. S., D. H. Peterson, P. R. Carlson, and T. J. Conomos. 1970. Some effects of freshwater inflow on the flushing of south San Francisco Bay: a preliminary report. *US Geol. Survey Circ.* 637-A.
17. Monismith, S. G., J. R. Burau, and M. Stacey. 1996. Stratification dynamics and gravitational circulation in northern San Francisco Bay. Pages 123-153 in J. T. Hollibaugh, editor. *San Francisco Bay: The Ecosystem*. AAAS, San Francisco.
18. Nixon, S.W., C.A. Oviatt, J. Frithsen, and B. Sullivan. 1986. Nutrients and the productivity of estuarine and coastal marine systems. *J. Limnol. Soc. S. Africa* 12:43-71.
19. Officer, C. B., T. J. Smayda, and R. Mann. 1982. Benthic filter feeding: A natural eutrophication control. *Marine Ecology Progress Series* 9:203-210.
20. Platt, T., 1978. Spectral analysis of spatial structure in phytoplankton populations in J.H. Steele (ed) *Patterns in Plankton Communities*, Plenum Press, NY 73-83.
21. Powell, T. M., J. E. Cloern, and L. M. Huzzey. 1989. Spatial and temporal variability in south San Francisco Bay (USA). I. Horizontal distributions of salinity, suspended sediments, and phytoplankton biomass and productivity. *Estuarine, Coastal, and Shelf Science* 28:583-597.
22. Powell, T.M., and A. Okubo, 1994. Turbulence, diffusion and patchiness in the sea. *Phil. Trans. R. Soc. Lond. B.*, 343:11-18.
23. Smith, S. V., and J. T. Hollibaugh. 2000. Water, salt, and nutrient exchanges in San Francisco Bay. *Interagency Ecological Program for the San Francisco Estuary*. Technical Report 66. Sacramento CA
24. Thompson, J.K., 1999, The effect of infaunal grazing on phytoplankton bloom development in South San Francisco Bay: PhD thesis, Stanford University, Stanford, CA, 419 p.
25. Torrence C., and G. P. Campo, 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1):61-78.

This page intentionally left blank

# CHARACTERIZATION OF FRACTAL STRUCTURES THROUGH A HAUSDORFF MEASURE BASED METHOD

FAHIMA NEKKA AND JUN LI

*Faculté de pharmacie et Centre de recherches mathématiques, Université de Montréal,  
C.P. 6128, Succ. Centre-ville, Montréal (Québec), CANADA H3C 3J7  
E-mail: fahima.nekka@umontreal.ca, li@crm.umontreal.ca*

We introduce a new method based on Hausdorff Measure Spectrum Function (HMSF) which provides a more precise way for tracing the geometrical organization of a fractal set. By way of its construction, this spectrum collects a rich amount of information that can be explored. A one by one step method is developed and applied to five Cantor-type sets to illustrate the ability of our method to distinguish between fractals having the same dimension. This results in a more complete identification of fractals. We also include some suggestive algorithms to construct the HMSF.

## 1 Introduction

Various physical processes and structures share the same fractal dimension in spite of their different appearance. The most used and popularized concept in fractal geometry was the fractal dimension. Only in the last decade, additional tools to deal with the degeneracy character of fractal dimension have been developed. Among them, few have been devoted to fractal features of 'texture', a broad concept called lacunarity by Mandelbrot <sup>1</sup>. It is largely documented that there is a need to develop methods that account for fine structure <sup>2,3</sup> since most of the properties depend, in addition to the fractal dimensionality, on other geometrical factors related to texture. To mathematically quantify this loose notion of texture, several methods have been proposed. The gliding-box algorithm (GBA) is one of them <sup>4</sup>. GBA has been derived from the box-counting method (BCM) by gliding a box over the set, one unit at a time in a discrete manner. Despite the popularity of this algorithm, it still proved to be degenerate in rather simple cases <sup>5</sup>, where two deterministic regular shapes of the same fractal dimension have not been resolved. In this fundamental context of characterization of the fine structure of a fractal, one would like to know if there is a way to build a more powerful quantitative and characteristic tool to unveil the information carried by the structure. Having this purpose in mind, we have been led to study the Hausdorff measure of the intersection of sets of same dimension with their translates <sup>6</sup>. The method we propose here is based on the Hausdorff measure of the translation of the set through itself in a continuous manner. Since the translation is made continuously on each point (local) and that Hausdorff measure (global) is estimated, the measure function obtained is able to extract the whole information within the structure. At this point, it should be mentioned that the indicator function of the intersection of a set with its translates can be viewed as a two-point joint moment (autocovariance) within the set's indicator function. This explains in a way why the measure function introduced here naturally completes the information obtained from pointwise descriptors. To explain the underlying mechanism and test our method, we will use Cantor-like

sets of same fractal dimension as a tractable model.

In this paper, we briefly recall our results concerning the properties of HMSF for the special case of triadic Cantor set. Then we extend this investigation for the larger family of uniform Cantor sets. We apply the developed method to typical Cantor sets having the same Hausdorff dimension. Finally, we propose an algorithm to numerically construct the HMSF.

## 2 Hausdorff Measure Spectrum Functions of Uniform Cantor Sets.

If  $C$  is the triadic Cantor set and  $C + t$  is the translate with the shift element  $t$ , it is known that they all have unit Hausdorff measure and dimension  $s = \log 2 / \log 3$ . To analyze the intersection  $I(t) = C \cap (C + t)$ , it is natural, as a subset of  $C$ , to study its Hausdorff measure at this dimension  $s$ . In our first paper investigating properties of intersection of Cantor set with its translates<sup>6</sup>, we have proved that for the triadic Cantor set  $C$ , the Hausdorff measure of  $C \cap (C + t)$  is not continuous in  $t$ . This measure can only take one of the values  $1/2^i$ ,  $i = 1, 2, \dots$ , when  $t$  has a finite triadic expansion; and it is of zero Hausdorff measure when  $t$  can not be expressed in a finite triadic expansion. Hence, this set of values forms a discrete spectrum of measures. Moreover, we also determined the exact expression of those elements  $t$  belonging to  $T_n = \{t : \mathcal{H}^s(I(t)) = 1/2^n\}$  where  $\mathcal{H}^s$  denotes the Hausdorff measure at dimension  $s = \log 2 / \log 3$ , that give rise to a particular measure  $1/2^n$  from the spectrum<sup>7</sup>. We showed that the elements of  $T_n$  can be grouped into an infinite tree structure where the number of branches to any knot of the tree is infinite and that they are given by

$$T_n = \bigcup_{\substack{1 \leq i_1, \dots, i_n < +\infty \\ s_1, \dots, s_n \in \{0, 1\}}} \{(-1)^{s_1} \frac{2}{3^{i_1}} + (-1)^{s_2} \frac{2}{3^{i_1+i_2}} + \dots + (-1)^{s_n} \frac{2}{3^{i_1+i_2+\dots+i_n}}\}. \quad (1)$$

## 3 Applications

In the following, we show that the discreteness of Hausdorff measure spectrum can be proved for a larger family of Cantor sets, namely for uniform Cantor sets and we give the exact expression of this Hausdorff measure spectrum.

First we recall the construction of the uniform Cantor set  $F$ .

Start from the unit interval  $[0, 1]$  that we denote as  $F_0$ .

In the first step, we replace  $F_0$  by  $m$  equally spaced subintervals of length  $r$ , the end points of  $F_0$  coinciding with the end points of the extreme subintervals. The formed set is denoted by  $F_1$ . We may write  $F_1 = \bigcup_{i=1}^m I_i$  where  $I_i$  are subintervals arranged from left to right in an increasing order.

Recursively, if  $F_{k-1}$  is composed of  $m^{k-1}$  subintervals of length  $r^{k-1}$  such that  $F_{k-1} = \bigcup_{i_1, i_2, \dots, i_{k-1}=1}^m I_{i_1, i_2, \dots, i_{k-1}}$  and  $|I_{i_1, i_2, \dots, i_{k-1}}| = r^{k-1}$ ,  $F_k$  can be obtained by replacing each subinterval  $I_{i_1, i_2, \dots, i_{k-1}}$  of  $F_{k-1}$  by  $m$  equally spaced subintervals of length  $r^k$ , the ends of  $I_{i_1, i_2, \dots, i_{k-1}}$  coinciding with the end points of the extreme subintervals. We may write  $F_k = \bigcup_{i_1, i_2, \dots, i_k=1}^m I_{i_1, i_2, \dots, i_k}$ ,  $|I_{i_1, i_2, \dots, i_k}| = r^k$ , and  $I_{i_1, i_2, \dots, i_{k-1}i} \subset I_{i_1, i_2, \dots, i_{k-1}}$  arranged from left to right in an increasing order.

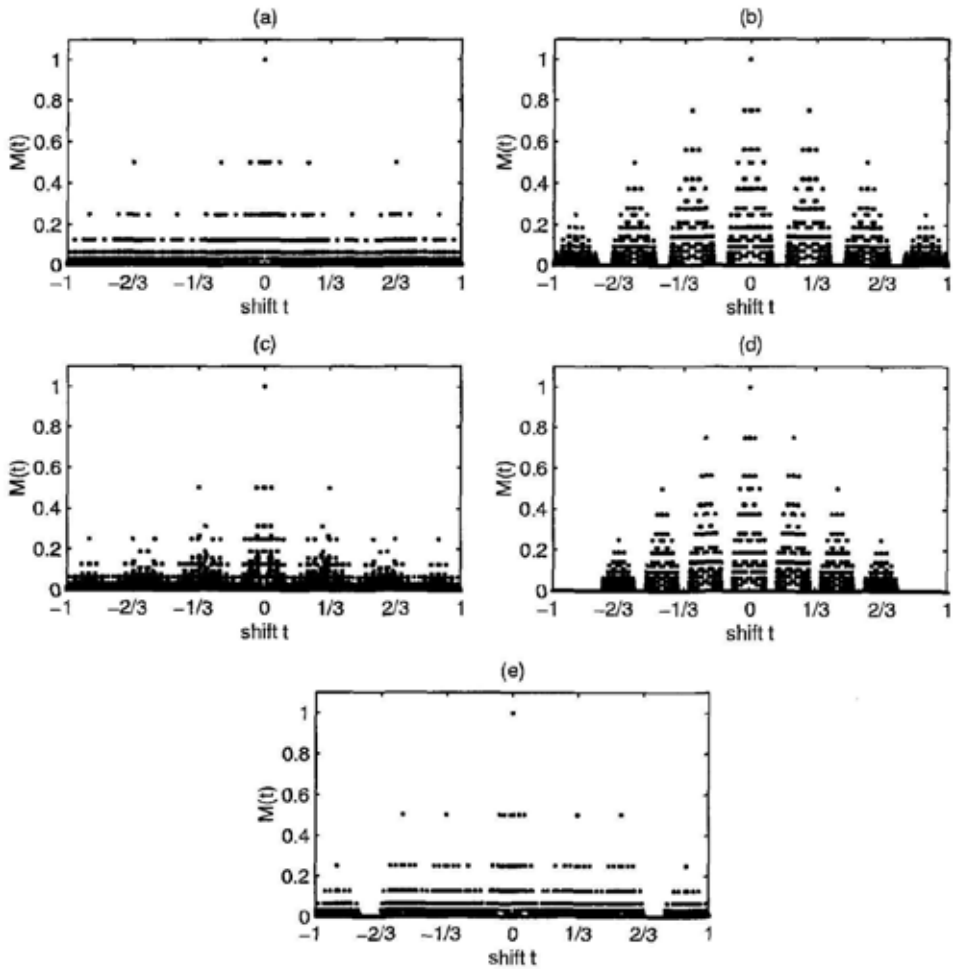


Figure 1. Hausdorff Measure Spectrum Functions for four Cantor-type sets that all have the same fractal dimension  $s = \log 2 / \log 3$ .

Naturally, the uniform Cantor set can be defined as the intersection of all  $F_k$ , i.e.,  $F = \lim_{k \rightarrow \infty} \bigcap_{k=1}^{\infty} F_k$ . For such a uniform Cantor set  $F = F(r, m)$ , if  $(2m-1)r < 1$ , then the Hausdorff measure at dimension  $s = -\log m / \log r$  of  $F \cap (F+t)$ ,  $-1 \leq t \leq 1$ , if it is not zero, can only take a value of the form

$$\left(\frac{1}{m}\right)^{i_1} \left(\frac{2}{m}\right)^{i_2} \dots \left(\frac{m-1}{m}\right)^{i_{m-1}}.$$

This can be seen in the following way: note that  $F \cap (F+t) = \bigcap_{k=1}^{\infty} (F_k \cap (F_k+t))$  where  $F_k \cap (F_k+t) \subset F_{k-1} \cap (F_{k-1}+t)$ . One can see that  $F_1 \cap (F_1+t)$  is composed of several subintervals (the case of isolated points is not considered), and since



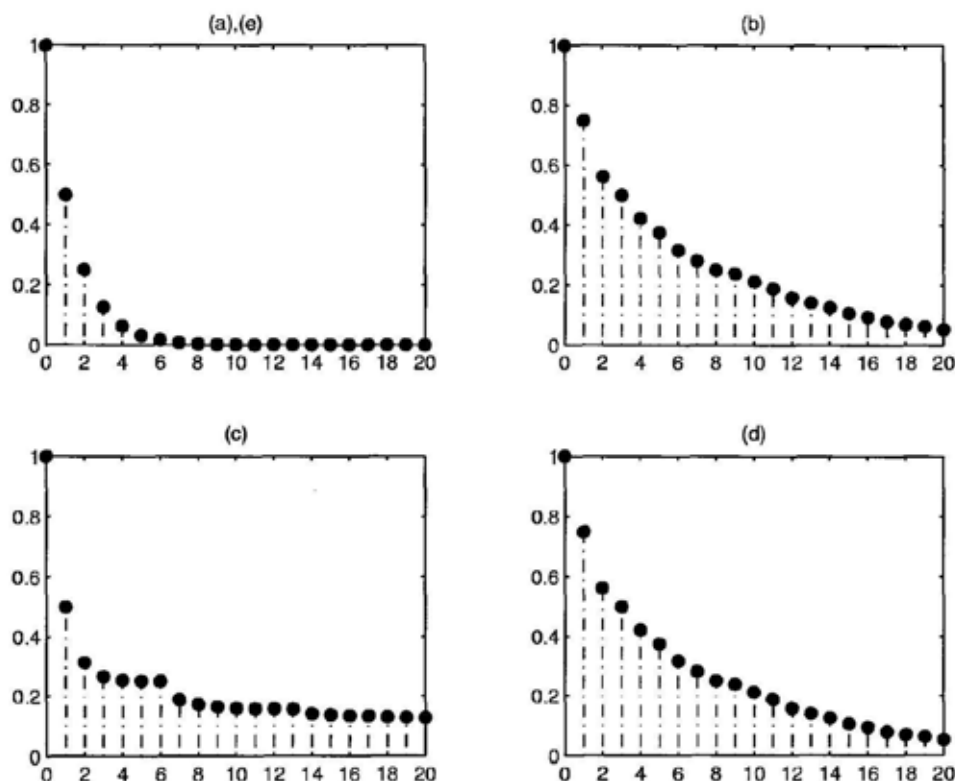


Figure 2. Hausdorff measure levels where the translational invariance of the Hausdorff measure is presented. (a), (b), (c), (d), and (e) refer to different fractal sets. The horizontal axis represents the level number.

$(2m-1)r < 1$ , this number is not larger than  $m$ . We find that the same number of subintervals of  $F_1$  is involved.

Assume that  $F_{k-1} \cap (F_{k-1} + t)$  is composed of several subintervals (the case of isolated points is not considered). Then, the same number of subintervals of  $F_{k-1}$  will be involved. Suppose that this number is  $1^{i'_1} 2^{i'_2} \dots m^{i'_m}$  where  $i'_1 + i'_2 + \dots + i'_m = k-1$ . As  $(2m-1)r < 1$ , a subinterval of  $F_{k-1}$  can only at most have no trivial intersection with one subinterval of  $F_{k-1} + t$ . The number of these intersection subinterval pairs from  $F_{k-1}$  and  $F_{k-1} + t$  is  $1^{i'_1} 2^{i'_2} \dots m^{i'_m}$ , all these pairs have the same intersection form. If each intersection form involves  $l : 1 \leq l \leq m$  subintervals of  $F_k$ , then the involved number of the subintervals  $I_{i_1, i_2, \dots, i_k}$  of  $F_k$  in  $F_k \cap (F_k + t)$  is  $1^{i'_1} 2^{i'_2} \dots l^{i'_l+1} \dots m^{i'_m}$ . If we rewrite  $i_j = i'_j$  for  $j \neq l$  and  $i_l = i'_l + 1$ , we find that this number is of the form  $1^{i_1} 2^{i_2} \dots l^{i_l} \dots m^{i_m}$  and  $i_1 + i_2 + \dots + i_m = k$ .

• If for a certain  $k$ ,  $F_k \cap (F_k + t)$  is composed of several subintervals of  $F_k$ , then,  $F \cap (F + t)$  is composed of  $1^{i_1} 2^{i_2} \dots m^{i_m}$  small uniform Cantor sets ( $F$  as prototype) of Hausdorff measure  $1/m^k$ . If we use  $\mathcal{H}^s$  to denote the Hausdorff

measure at dimension  $s$ , then

$$\mathcal{H}^s(F \cap (F+t)) = \left(\frac{1}{m}\right)^{i_1} \left(\frac{2}{m}\right)^{i_2} \dots \left(\frac{m-1}{m}\right)^{i_{m-1}}.$$

• If for any  $k$ ,  $F_k \cap (F_k+t)$  is not composed of subintervals of  $F_k$ , then, for any  $k = i_1 + i_2 + \dots + i_m$ ,

$$\mathcal{H}^s(F \cap (F+t)) \leq \left(\frac{1}{m}\right)^{i_1} \left(\frac{2}{m}\right)^{i_2} \dots \left(\frac{m-1}{m}\right)^{i_{m-1}}.$$

$F \cap (F+t)$  is in the union of several small uniform Cantor sets ( $F$  as prototype) of Hausdorff measure  $1/m^k$  and the number of these uniform Cantor sets is the same as the number of the corresponding subintervals of  $F_k$  involved in  $F_k \cap (F_k+t)$ .

Since the right term of the above inequality goes to zero when  $k \rightarrow +\infty$ , then

$$\mathcal{H}^s(F \cap (F+t)) = 0.$$

In the following, we apply our method to differentiate between sets having the same fractal dimension ( $\log 2/\log 3$  in this example). These sets are constructed from the initiator  $I = [0, 1]$  and their respective generators are defined by the following iterative function systems:

- (a)  $S_i : x \mapsto \frac{1}{3}x + \frac{2(i-1)}{3}$  for  $i = 1, 2$ ;
- (b)  $S_i : x \mapsto \frac{1}{9}x + \frac{8(i-1)}{27}$  for  $i = 1, 2, 3, 4$ ;
- (c)  $S_i : x \mapsto \frac{1}{9}x + \frac{2(i-1)}{9}$  for  $i = 1, 2$ ,  $S_i : x \mapsto \frac{1}{9}x + \frac{3i-4}{9}$  for  $i = 3, 4$ ;
- (d)  $S_i : x \mapsto \frac{1}{9}x + \frac{2(i-1)}{9}$  for  $i = 1, 2, 3, 4$ ;
- (e)  $S_1 : x \mapsto \frac{1}{9}x$ ,  $S_i : x \mapsto \frac{1}{9}x + \frac{2i-1}{9}$  for  $i = 2, 3$ , and  $S_4 : x \mapsto \frac{1}{9}x + \frac{8}{9}$ .

The HMSF of the above fractal sets are shown in Figure 1, 3. We propose here two different ways to exploit the HMSF in order to distinguish between these sets. With the translation invariance based method (TIBM), we take the translation invariant values of HMSF corresponding to values preserved by translation. Each value corresponds to a level, which is a set of points representing a fixed HMSF value for different shifts (see Figure 1). The graph of these levels, in terms of the shift number, are illustrated in Figure 2 (all the measures have been normalized). We see that TIBM succeeds in distinguishing between (a), (c) and (b)(or (d)). However, TIBM levels are the same for (a) and (e) as well as for (b) and (d). This last fact does not allow one to conclude that (a) and (e) or (b) and (d) are the same. We have yet to go a further step in our exploration and use the fixed level based method (FLBM). This method compares, for a given level, the HMSF values of the concerned sets. In fact, the first level, which contains the whole information of HMSF, is enough. In figure 3, we plotted the first four fixed levels (from 0 to 3) of the HMSF of (a) and (e). Graphically, the difference is already obvious on level one. This difference can be quantified by averaging weighted distances between shift values and the shift accumulation point at the first level, giving thus

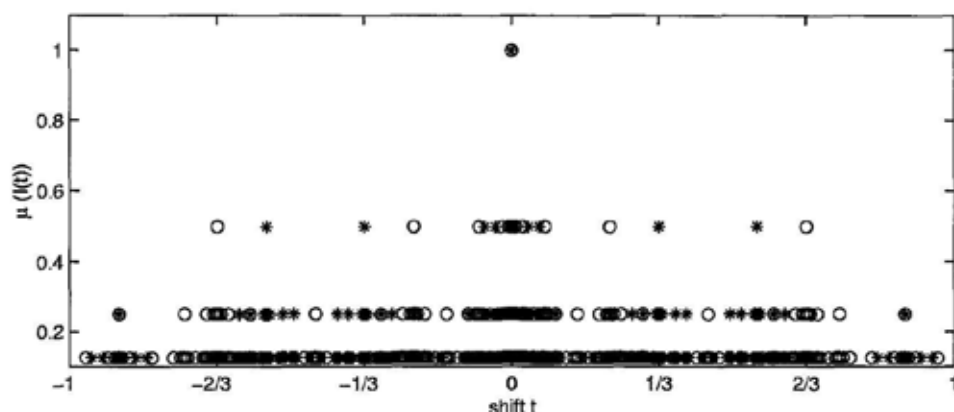


Figure 3. The first four levels of Hausdorff measure function, represented by (o) and (\*) for the Cantor set (a) and (e) respectively.

level indexes associated to each set. For example, using a dyadic sequence weights  $\{1/2^i\}_i$ , we get the value 0.5962 for (a) and 0.6248 for (e). This index, from one part, is able to differentiate between sets and, from the other part, indicates the degree of homogeneity of the set: the higher the index, the more homogeneous is the set. Finally, for sets (b) and (d), one observes that they have the same HMSF once the support corresponding to (d) is reported, by a rescaling centered on 0, on the same support of (b) (see Figure 1). This bilateral matching is naturally also reflected in the FLMB that shows the same fractal structure between (b) and (d).

Now, we propose an algorithm to approximate the HMSF, which, in a concern of simplicity, is illustrated for the triadic Cantor set. Since this algorithm is solely based on the similarity properties, it can easily be applied for any fractal set.

This algorithm is based upon the similarity properties of a fractal set which are inherited by the HMSF itself. This similarity can be obtained by a simple check of the HMSF.

We always use  $C$  to represent the triadic Cantor set generated from the unit interval  $[0, 1]$  and consider the intersection of  $C$  with its translation  $C + t$ , which we denote by  $I(t) = C \cap (C + t)$ .  $t$  varies between  $-1$  and  $1$ .

Let  $C = C_l \cup C_r$ , where  $C_l$ ,  $C_r$  are respectively the identical left and right parts of  $C$ . Clearly,  $C_r = C_l + 2/3$  and we have

$$\begin{aligned} C \cap (C + t) &= (C_l \cup C_r) \cap ((C_l + t) \cup (C_r + t)) \\ &= [C_l \cap (C_l + t)] \cup [C_l \cap (C_r + t)] \cup [C_r \cap (C_l + t)] \cup [C_r \cap (C_r + t)]. \end{aligned}$$

If we use  $\mathcal{M}(t)$  to express the Hausdorff measure of the intersection  $I(t)$  for the right four terms, at Hausdorff dimension  $s = \log 2 / \log 3$ , we have

$$\mathcal{M}_{ll} = \mathcal{H}^s(C_l \cap (C_l + t)) = \frac{1}{2} \mathcal{M}(3t) \quad (2)$$

$$\mathcal{M}_{lr} = \mathcal{H}^s(C_l \cap (C_r + t)) = \frac{1}{2} \mathcal{M}(3(t + \frac{2}{3})) \quad (3)$$

$$\mathcal{M}_{rl} = \mathcal{H}^s(C_r \cap (C_l + t)) = \frac{1}{2} \mathcal{M}(3(t - \frac{2}{3})) \quad (4)$$

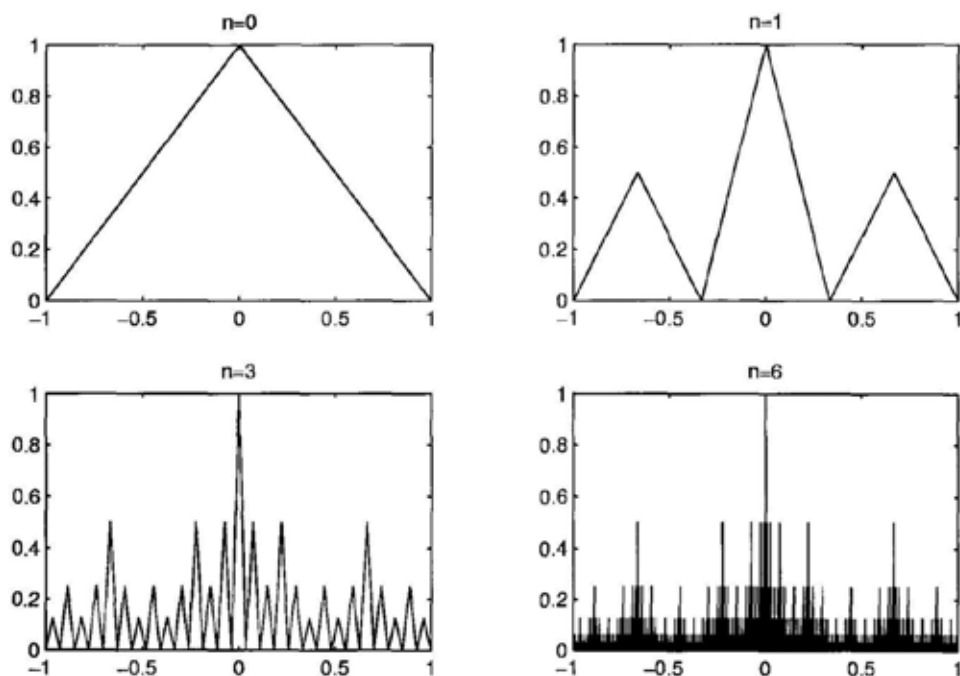


Figure 4. The approximation of the Hausdorff measure spectrum function by iteration, the generations: 0, 1, 3, 6.

$$\mathcal{M}_{rr} = \mathcal{H}^s(C_r \cap (C_r + t)) = \frac{1}{2} \mathcal{M}(3t) \quad (5)$$

Finally, we obtain the similarity equation verified by the HMSF of the triadic Cantor set  $C$ , *i.e.*

$$\mathcal{M}(t) = \frac{1}{2} \mathcal{M}\left(3\left(t - \frac{2}{3}\right)\right) + \mathcal{M}(3t) + \frac{1}{2} \mathcal{M}\left(3\left(t + \frac{2}{3}\right)\right). \quad (6)$$

If we denote by  $G$  the graph of  $\mathcal{M}(t)$ , it is easy to see that  $G$  is the attractor of three affine maps:  $S_0, S_1, S_2$ , *i.e.*

$$G = S_0(G) \cup S_1(G) \cup S_2(G)$$

where

$$\begin{aligned} S_0 \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}; \\ S_1 \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}; \\ S_2 \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{2}{3} \\ 0 \end{pmatrix}. \end{aligned}$$

The process of the similarity algorithm is shown in Figure 4.

We can also approximate the HMSF by a sequence of continuous functions, obtained by interpolation.

#### 4 Conclusion

In summary, this paper extends our results on intersection of triadic Cantor sets<sup>6,7</sup>, to a larger family of fractals. The classification ability of the method is illustrated on some typical examples having the same Hausdorff dimension. We exploited the discreteness of the HMSF of these examples to distinguish between them. Similar results on more general sets have been obtained and will be published elsewhere. Moreover, other potential applications of the HMSF within the fractal analysis context is in process.

#### Acknowledgments

This work has been supported by the Natural Sciences and Engineering Research Council grant hold by F. Nekka (NSERC RGPIN-227118).

#### References

1. B. B. Mandelbrot, *The Fractal Geometry of Nature*, (W.H. Freeman, San Francisco, 1983).
2. A. Arneodo et al, *Phys. Lett. A* **124**, no.8, 426 (1987).
3. Y. Gefen, A. Aharony and B. B. Mandelbrot, *J. Phys. A: Math. Gen.* **17**, 1277 (1984).
4. C. Allain and M. Cloitre, *Phys. Rev. A* **44**, 3552 (1991).
5. D. A. Fabio, A. Reis and R. Riera, *Phys. A: Math. Gen.*, **27**, 1827 (1994).
6. F. Nekka and J. Li, *Chaos, Solitons and Fractals* **13**, no.9, 1807 (2002).
7. J. Li and F. Nekka, *Chaos, Solitons and Fractals*, to appear.

# FRactal scattering indicators for urban sound diffusion

Philippe Wołoszyn

Cerma UMR CNRS 1563, Ecole d'Architecture de Nantes, BP 81931

F-44319 Nantes Cedex 3, France

E-mail : [philippe.wołoszyn@cerma.archi.fr](mailto:philippe.wołoszyn@cerma.archi.fr)

Irregular surfaces like urban facades produce an anomalous back-scattered region, creating an acoustic interference field in their neighborhood. Thus, in order to be able to detect that scattered energy's minima and maxima through taking the facade morphological characterization into account, we propose a new measurement method of the building geometry, using mathematical morphology techniques. Results of this geometrical approach provide two types of indicators, global and local. The global one, the structure factor of the urban facade, is related to the multiscale characterization of the whole building geometry through the computation of the spatial Fourier transform of the scatterers. The complementary local indicator evaluates the vertex multiscale densitometrical distribution at each incidence angle, provided through a fractal evaluation technique, the Minkowski sausage. This densitometry computation reveals the characteristic directions of scattering, which has to be calculated through the scattering pressure function along the lateral active diffraction zone.

## 1 Introduction: Problem and purpose

The exterior facade of a typical urban building does not reflect noise in a purely specular manner. Because the dimensions of the irregularities (decorative elements, windows, balconies,...) are comparable to the sound wavelengths, the major type of reflections on the buildings is scattering, inducing a global diffusion behaviour of sound in an urban street. Consequently, irregular surfaces like urban facades produce an anomalous back-scattered region, with the creation of an acoustic interference field in its neighborhood. Thus, in order to be able to detect that scattered energy's minima and maxima, we have to take into account both incidence angle and multiscale characterization for diffusive evaluation of urban surfaces through mathematical morphology techniques.

## 2 Diffusion through oblique incident wave

### 2.1 A first approximation: the Rayleigh criterion

Historically, the first attempt at determining the scattering amplitudes was made in 1893 by Lord Rayleigh, who assumed a unique solution for the wave equation for the whole boundary of a  $\Lambda$ -corrugated surface [1]. Concerning the inferior diffusion limit frequency, Rayleigh's work proposes a phase grating calculation between two acoustic rays related in [2], which takes into account the source incidence angle. Taking into account the path difference  $\Delta d$  between two rays with wavelength  $\lambda$  and incidence angle  $\alpha$  regarding a surface with depth  $\Lambda$  provides the following phase grating calculation between the two rays:

$$\Delta\alpha = \Delta d (2\pi / \lambda) = \cos \alpha (4 \pi \Lambda / \lambda) \quad (1)$$

with the path difference  $\Delta d = 2 \Lambda \cos \alpha$ . For a weak path difference  $\Delta d$ , rays are coherent and the acoustic wave is specularly reflected. Increasing  $\Delta d$  interferes with rays, till  $\Delta d = \pi$ , so that no energy is displayed in the specular direction : sound energy is diffused.

Rayleigh criterion defines the limit between specular and diffuse behaviour of an incident source, corresponding to the facade depth irregularities as :  $\Lambda < 1/8 \cos \alpha$ .

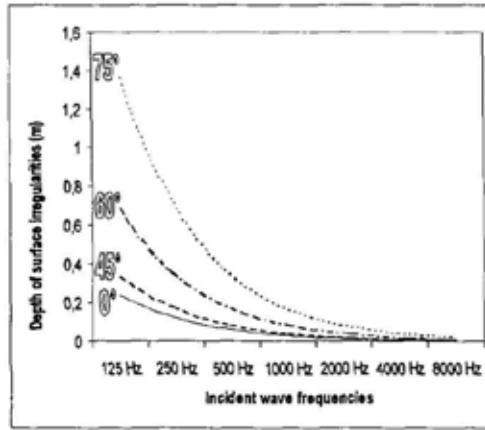


Figure 1. Rayleigh criterion: Quantification of the dimensional limit between specularity and diffusion, as a function of frequency and incidence angle.

The specular reflection zone is defined in the lower part of the facades depths irregularities curves, taking the frequency and the angle of the incident wave into account.

## 2.2 Diffraction densitometry of an indented plane

Following those non-specularity conditions, the propagation directions specified by the unit vectors  $\mathbf{v}(d) = (\phi_d, 0, \gamma_d)$  for a given regular plane division, repeating  $n$  times a spatial unit of width  $\Lambda$  are defined as the *characteristic directions of scattering* associated with the localization length  $\Lambda$ :

$$\sin \alpha_d = \frac{p\lambda}{2n\Lambda} - \sin \alpha, \quad (2)$$

where  $\alpha_d$  is the grazing diffraction angle made by vectors  $\mathbf{v}(d) = (\phi_d, 0, \gamma_d)$  in the direction  $Ox$ , with  $\phi_d = \cos \alpha_d$  and  $\gamma_d = \sin \alpha_d$ , and  $p$  the diffraction order. For  $\Lambda = 0$ , equality between incident and reflected angle remain true (specular conditions). We can note here that, ignoring the specular component  $\sin \alpha$ , the first term of the previous equation can be compared to Bragg's Law, leading to the following expression:

$$\sin \alpha = \frac{p\lambda}{2d}, \quad (3)$$

where the integer  $p$  is the diffraction order,  $d = n\Lambda$ , the distance between two reflection planes,  $\lambda$ , the wavelength of the incident beam and  $\alpha$  its incidence angle. This law, also used in the field of crystallographic diffractometry, gives the conditions for constructive interferences, which is producing strong diffraction. Through this equation, Sir W.H. Bragg and his son developed a simple semi-quantitative model to express the diffraction from 3-D crystals, explaining why the cleavage faces of crystals appear to reflect X-rays beams at certain angles of incidence. Considering those crystal structures as families of parallel planes ( $hx, ky, lz$ ) running in different directions, each plane acts like a slightly reflective mirror, reflecting a tiny fraction of the incident beam. When in phase, those reflections lead to constructive interferences, conditioned with Bragg's Law equation. For

$p=1$ , all planes inside the cosine scatter in phase, providing maximal diffraction. For  $p=1-\Lambda$ , the diffraction cancels.

In the same way, Bragg's law and Rayleigh criterion defines the conditions of interfering behaviour, for acoustic waves reflected on micrometer or meter-scaled parallel planes. Consequently, the pure diffracted energy part of an urban indented surface can be expressed through Bragg's Law, considering the path difference between two facades surfaces, as between two crystal planes, for the constructive interference conditions:  $p\lambda=2n\Lambda \sin\alpha$ , where  $n\Lambda$  is the facade indentation depth. Moreover, the two-dimensional polar response of a given indented surface can be expressed through the diffraction orders ( $p, q$ ), taking the angles of incidence and diffraction into account:

$$\sqrt{p^2 + q^2} = n\Lambda \frac{\sin \alpha_d + \sin \alpha}{\lambda} \quad (4)$$

The characteristic directions are those along which the waves emanating from the individual facade indentation depth  $\Lambda$  are exactly in phase. This constructive interference condition is both conditioned with the adimensional modulus  $\Lambda/\lambda$ , which quantifies the energy of non-evanescent scattering losses, represented by the area of scattering intensity pattern lobes, and with the previous Bragg's equation  $p\lambda=2n\Lambda \sin\alpha$ . Indeed, this modulus is conditioning the solutions of equation (4), as  $\Lambda/\lambda = 0$  confirms specular reflection conditions (as the diffraction order  $p$  is null through the limit of the diffusion), and as this modulus value conditions its number of real solutions, corresponding to the diffraction directions (lobes of the surface's radiated energy).

### 3 Diffraction and structure factor of a multiscale rough boundary surface

For all other directions, the reflected waves will destructively interfere, resulting in complete cancellation for a self-similar periodic structure. For non- or pre-fractal structures as urban facades, the scattered field will show mainlobes in the characteristic directions, and sidelobes elsewhere for a given sound frequency).

#### 3.1 Phase of diffraction

The interference conditions can be expressed also by defining the phases of the incident wave vector  $\mathbf{k}_0$  and the diffracted vector  $\mathbf{k}$ , which both have an amplitude equal to the reciprocal of the wavelength. In order to calculate the phase of the diffracted wave, taking the path lengths difference  $\Delta d = 2 n\Lambda \cos\alpha$  into account, we will consider the difference between the path of the sonic particle (phonon) along the incident beam  $\mathbf{k}_0\mathbf{r}$  and its path along the diffracted beam  $\mathbf{k}\mathbf{r}$ . By expressing this path length difference  $\Delta d = 2 n\Lambda \cos\alpha = \mathbf{k}_0\mathbf{r} - \mathbf{k}\mathbf{r}$ , the overall diffraction phase will be written as  $-2\pi(\mathbf{k}_0\mathbf{r} - \mathbf{k}\mathbf{r}) = 2\pi(\mathbf{k} - \mathbf{k}_0)\mathbf{r}$  (figure 2).



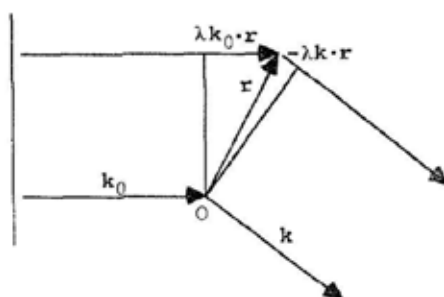


Figure 2. Phase geometry

Considering  $|r| \cos \alpha$ , the component of  $r$  in the direction of the diffraction vector  $s$ , all points with the same value of  $sr$  are lying on a plane perpendicular to vector  $s$ , allowing the same diffraction phase (figure 3).

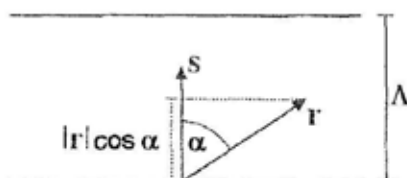


Figure 3. Diffraction geometry

Consequently, as the length of the diffraction vector  $|s|$  is equal to  $1/\Lambda$  (inverse of the indentation depth),  $sr$  is equal to the distance between two Bragg planes (or indentation surfaces), and diffraction from any point  $r$  will have a phase of  $2\pi sr$ . Moreover, we can define the phonon density resolution through the mean resolution distance, according to the period  $n\Lambda$  of the Fourier series of the phonon density map. The following equation leads us to the calculus of the reflective resolution of the structure, involving the path difference  $\Delta d$  as:

$$\frac{1}{\Delta d} = \sqrt{\frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}} = \frac{1}{2\Lambda \cos \alpha} \quad (5)$$

where  $(h, k, l)$ , the Miller indices, specify the direction and the period of the tridimensional cosine wave  $\cos(2\pi/[hx/a + ky/b + lz/c])$ .

### 3.2 Structure factor indications: spatial scattering function

When measuring several phonons located at different points, the diffraction at each point will be the sum of the waves scattered by each phonon. So, the expression of this sum with Euler's equation gives us, for the  $j^{\text{th}}$  phonon with coordinates  $(x_j, y_j, z_j)$ :

$$F(s) = \sum_j \exp(2\pi i s \cdot r_j) \quad (6)$$

This wave is represented here by its *structure factor*, which is the Fourier transform of the scatterers of equal strength on all points of the diffraction plane. Continuous expression of this previous equation involves the phonon's density  $\rho(r)$  as:

$$F(s) = \int_{\text{space}} \rho(r) \exp(2\pi i s \cdot r) dr \quad (7)$$

As shown through this expression of the structure factor, the diffraction pattern is defined as the Fourier transform of the phonon density.

Taking into account the tridimensional distribution of phonons into the diffusive structure involving Miller indices ( $h, k, l$ ), as the plane perpendicular to vector  $s$  can be written as:  $\mathbf{s}r_j = hx_j + ky_j + lz_j$ , we can afford the previous spatial expression of the structure factor in the three dimensions of space with integrating the tridimensional cosine wave as:

$$F_{hkl} = \iiint_{\text{Phonon Volume}} \rho(x, y, z) \exp(2\pi i(hx + ky + lz)) dx dy dz \quad (8)$$

The acoustic field is then expressed through the phonon's density  $\rho(x, y, z)$ , which is useful in calculation of the mean square diffracted sound pressure  $P_d$  by the whole volume for a given distance of the structure [3]:

$$P_d^2 = \rho \iiint_V \frac{\cos \alpha \cos \alpha_d}{\pi r_0^2} P^2 dV \quad (9)$$

where  $r_0$  represents the distance from the receiver to the structure, and  $P$  the incident wave pressure. One can remark that the function of the cosine of the angle between the direction of observation and the normal to the surface in the observation point reminds us of the Lamberts law [4], which is assumed to represent the physical behaviour of sound or light after reflection on an ideally diffusing surface. As mentioned previously, the angular repartition of the sound energy is computed with the Miller indices, involved through the individual phase contributions  $2\pi i(hx_j + ky_j + lz_j)$ , which represents the *spatial scattering function* of the reference volume  $V$  (equation (8)).

### 3.3 Dynamic scattering function

This leads us to consider the structure factor as a function of time, called *dynamic structure factor*, or *dynamic scattering function*, by introducing time  $t$  through a random walk in random environment [5]:

$$F(\mathbf{s}, \omega) = 2 \int_0^\infty \cos(\omega, t) \rho(\mathbf{r}, t) dt \quad (10)$$

The dynamical density distribution  $\rho(\mathbf{r}, t)$  can be obtained with the probability for a sonic particle to walk to location  $\mathbf{r}$  during the time  $t$   $P(\mathbf{r}, t)$ , that remains equation (5), with the following relationship [6]:

$$\rho(\mathbf{s}, i\omega) = \int d\mathbf{r} \exp(i\mathbf{s}\mathbf{r}) \int_0^\infty dt \exp(-i\omega t) P(\mathbf{r}, t) \quad (11)$$

with  $P(\mathbf{r}, t)$ , describing the sonic particle's probability for a *fractional Brownian walk* in a non-integer (fractal)  $D$ -dimensional space [7], [8]:

$$P(\mathbf{r}, t) \rightarrow \frac{1}{(4\pi\delta t)^{D/2}} \exp \frac{r^2}{4\delta t}, N \rightarrow \infty \quad (12)$$

where  $\delta$  is the diffusion coefficient of the  $D$ -dimensional structure.

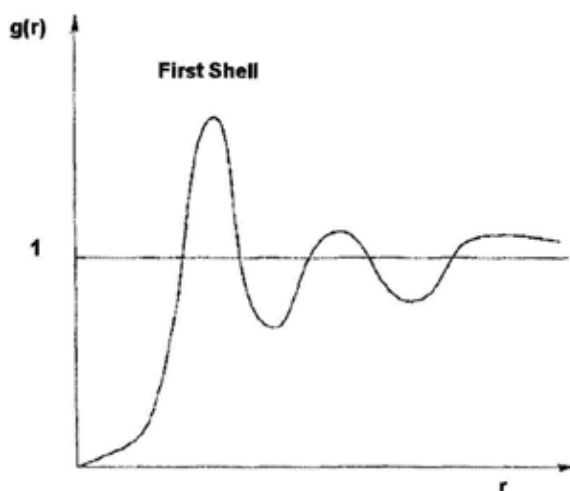
### 3.4 Angular distribution function

This function can be evaluated using the Laplace transform [9]. After integrating over the angles, we obtain, for the  $p_d^{\text{th}}$  diffraction order:

$$\rho(s, i\omega) = \frac{2\pi g}{i\omega^{1-D/d_w}} \int dr r^{D-1} \frac{\sin(p_d r)}{p_d r} * \frac{1}{(i\omega r^{1/d_w})^{D-1/2}} \exp(-i\omega r^{1/d_w} / g) \quad (13)$$

where  $D$  is the fractal dimension of the diffusive structure,  $d_w$  the dimension of the random walk of the phonon [10], and  $g$  is the *angular distribution function* along a characteristic direction of scattering:

$$g(r) = \frac{1}{m^2} (G_0(r) - \bar{m} \rho(r)) \quad (14)$$



**Figure 4.** Typical angular distribution function. The first shell represents the main density function at a distance  $r$  of the structure.

$G_0 = \langle m(o)m(r) \rangle$  is the second moment of density taken in the points  $0$  and  $r$ —called *density-density correlation function*. Square of the average density  $\bar{m} = \sum_i \rho(r_i)$  constitutes the limit of the density-density correlation function  $G_0$  when  $r \rightarrow \infty$ :  $G_0(r) \rightarrow \bar{m}^2$  and  $g(r) \rightarrow 1$ . This function defines the scattering intensity of the structure for a defined angle as:

$$I_q = \int_j G_0(r)^{-iqr} dr \quad (15)$$

### 3.5 Parseval theorem and diffusion volume

Parseval's theorem formulates that the energy in the frequency domain is the same as the energy in the spatial domain [11]. Consequently, the mean square value on one side of the Fourier transform equation (8) is proportional to the mean square value on the other side. So Parseval's theorem does allow to express the phonons density distribution as a transform of the spatial distribution of the surface scatterers. This property allows us to express the angular distribution function as a discrete quadratic summation of elementary structure factors as following:

$$\int_V \rho(x, y, z)^2 dV = \frac{1}{V} \sum_{xyz} |F_{hkl}|^2, \quad (16)$$

where the diffusion volume  $V_{xyz}$  is a ratio between the square root of the discrete quadratic summation of the structure factors and the angular distribution function. This diffusion volume  $V_{xyz}$  is experimentally obtained by a mathematical morphological measure using a Minkowski operator, which provides a ribbon surface constituting a neighborhood area, under the condition of continuity [12].

Considering the Minkowski analysis of a tridimensional structure, analyzed with a structuring element of variable radius  $\Lambda$ , the phonons density-density distribution can be expressed through the roughness autocorrelation of the diffusive structure, involving the diffusion volume  $V_{xyz}$ , and the structure factor  $F_{hkl}$ , which defines both the global and local behavior of the structure as:

$$\rho(x, y, z) = \sqrt{\frac{\sum_0^{\Lambda_{max}} |F_{hkl}|^2}{V_{xyz}}} \propto V_{xyz} \sum_0^{\Lambda_{max}} \sin \alpha_d + \sin \alpha \quad (17)$$

#### 4 Application : a Facade scattering characterization

##### 4.1 The urban facade model

The spatial configuration we measure here is a numerical 3-D model of a neoclassical facades of an urban street of Nantes, France, the *rue d'Orléans*, belonging to a 19<sup>th</sup> urban morphology type, with windows, doors, and freestone casting. One of the main characteristics of this type of architecture is the relative exuberance of its facade structure, following neo-classical composition. This facade is considered as a tridimensional object situated in an ortho-normal space, rotating around the Z-axis (figure 5):

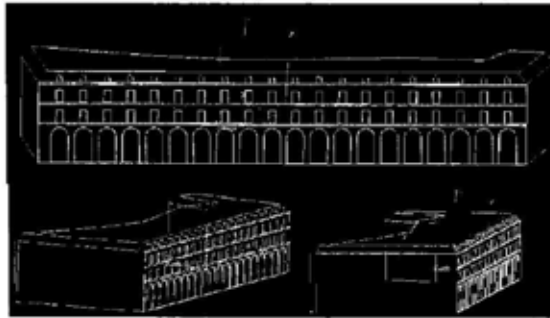


Figure 5. Orthographical rotative analyze of a facade in the *rue d'Orléans*, Nantes.

##### 4.2 The Minkowski measurement technique

In order to characterize the scattering behavior of the volume of the facade, we apply a fractal Minkowski operator, called Minkowski sausage, to evaluate the vertex multiscale densitometry distribution at each incidence angle. This operator replaces each point of the vertexes of the urban geometry with a sphere with variable radius  $\Lambda$ , as seen figure 6:

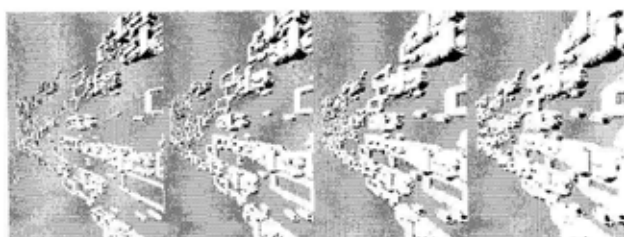


Figure 6. Tridimensional dilation of the urban scene vertexes (perspective views).

This transformation of the urban geometry corresponds to the *dilation* operation in the morpho-mathematical context [12]. The union of all spheres is called the *3-D Minkowski sausage*. The variation of their diameter gives us successive perimeter/surface ratios at each viewing angle, and regression evaluates the fractal dimension of the structure, in a specified validity domain.

#### 4.3 Facade fractal measurement

The spatial multiscale evolution of the perimeter-surface ratio  $P/S$  defines the profile's *Shape spectrum* of the facade [13]. This spectrum defines the multiscale relationship between the radius evolution of the spherical structuring element and the "mass" of the structure, for each of the angular measure. As seen in the following figure, the specular domain is illustrated by a strong decrease of the  $P/S$  ratio, which corresponds to the limit  $\Lambda_{max}$  for the radius of the structuring element. For this domain, the Euclidean dimension  $d$  and the fractal dimension  $D$  of the mean structure reach the same value. This break in the facade indentations shape spectrum behavior occurs for  $\Lambda_{max} = 100$  cm, for every value of incidence angle  $\alpha$  (figure 7):

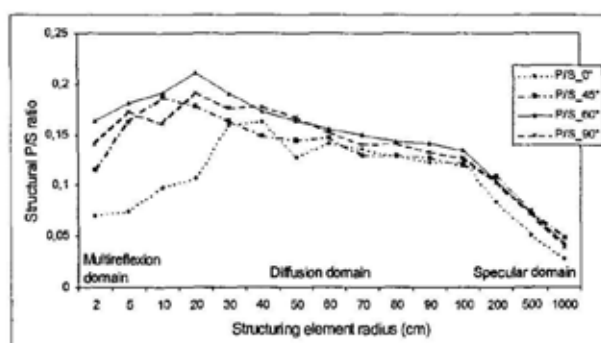


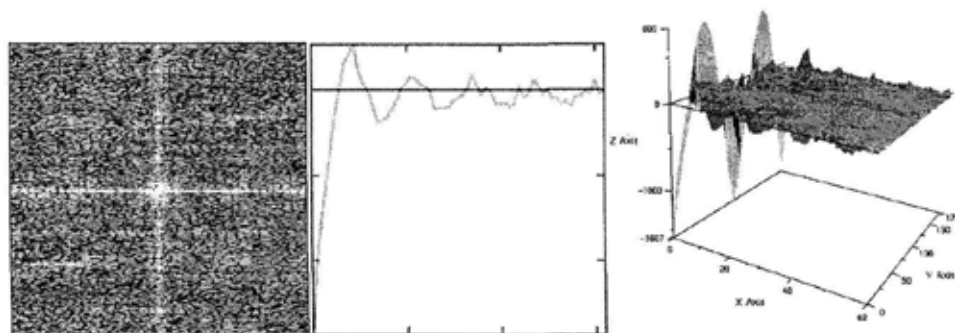
Figure 7. Urban facade shape spectrums using a spherical recovering element at different incident angles.

#### 4.4 Results of the analysis : Facade's structure factors and vertex densitometry

The Fourier transform of the surface roughness reflects the facade's complexity, leading both to the angular distribution function and the spatial scattering function of the indented surface calculations [14].

As an indicator of the indentation frequency, the Fourier transform discriminates clearly the structure of a surface, revealing the spatial occurrences of the roughness peaks (figure 8.1).

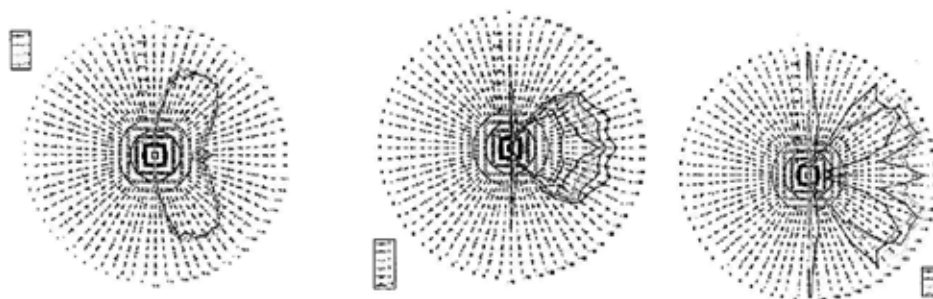
The angular distribution function is defined through the structure factor computation of the surface, and indicates the facade scattering behaviour for a particular direction of the incident beam (figure 8.2). Moreover, the spatial scattering function offers a tridimensional interpretation of the angular distribution function, showing the distribution of the surface's scatterers along every incidence angle of the acoustic source (figure 8.3).



**Figure 8.** Fourier transform (1), angular distribution function (2) and spatial scattering function (3) of the facade indented surface.

These indicators allow the computation of the vertex densitometry for every incidence angle of the surface, with an increment of 5 degrees for localisation lengths  $\lambda$  from 0.05 to 10 m, which corresponds to acoustical frequency domain of between 25 Hz and 8 kHz. These densitometries correspond to the characteristic directions of scattering, through the calculation of the density distribution  $\rho(x, y, z)$  for each incidence angle.

This angular evaluation of the vertex distribution shows azimuthal densitometries due to interreflexions of the corners and the freestone casting along three windows depth, corresponding to the lateral active diffraction zone.



**Figure 9.** Angular vertex densitometry of the urban facade applied to the urban scene. Measures for roughness values  $r$  of, respectively, 1 to 10 m, 0.2 to 0.5 m and 0.05 to 0.2 m.

Global polar responses for growing localisation lengths show globally a decreasing diffusivity, revealing a bilobe distribution structure of the biggest scatterers, a cardioid for middle-sized ones and very characteristic peaks for high frequency roughness.

## 5 Experimental validation: *In situ* measurements

In order to validate this geometrical scattering characterization model, we attempt to define a new method of measurement, applied *in situ* on a neoclassical facade in the

historical heart of Nantes, the *Hôtel Deurbrouck*, with windows, free-stone casting, and delicate metal guardrail and railing.

### 5.1 Measurement Methodology

Fractal scattering characterisation is here compared to experimental results at each incidence angle, by using a measurement procedure exploiting maximum-length sequences stimulus (MLS). That means a binary value pseudo random sequence with a period :  $L = 2^N - 1$  is generated by a N level digital register. The MLS have a quasi flat spectrum, and autocorrelation provides a Dirac function. Moreover, its significant signal / noise ratio avoids high peak factor and enables urban background noise decorrelation. The measurement procedure consists in emitting this MLS impulse sound wave, and then recording incident wave (time t) and the diffractive wave (time t+dt) of the back-scattered region of the facade. The resulting signal, constituting the facade impulse response, is analyzed to pull the incident wave away from the rest of the signal by time windowing. The content of this window is then analyzed in frequency domain, applying the Fourier transform, and provides the facade's Frequency Transfer Function. Several positions of source and microphone provide the reflection law of the facade.

### 5.2 Measurement instrumentation system

This manipulation involves a Supravox broadband 215RTF-bicone 21cm loudspeaker, enclosed with a  $\frac{1}{4}$  length wave system, installed at 6.8m height in the front of the building on a mobile system. The 1/2-inch B&K microphone has a similar height and can move on a half-circle with step of  $5^\circ$ , using a laser pointer for precise setup, in order to vary the reception incidence angle step by step. Facade impulse response is then recorded for each angle, through a microcomputer equipped with a numerical acquisition card and software emitting MLS, and processing data (MLSSA V.10.0). The microphonic post-treatment data provides the back-scattered reflection law for each frequency. Subtraction between the resulting signal and the same measurement method applied on a specular smooth urban surface provides the scattering effect of the facade extrusions.

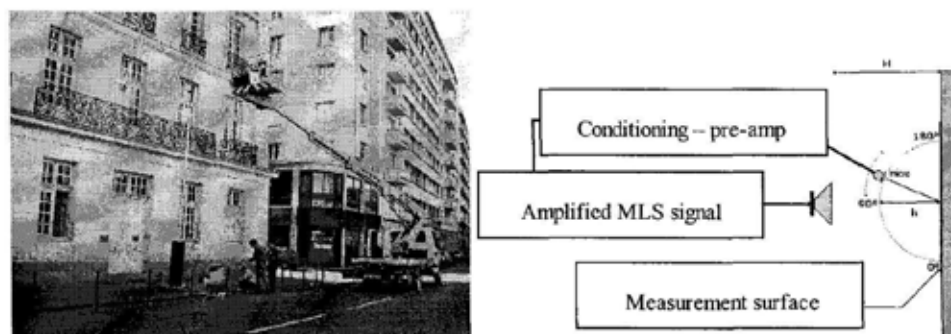


Figure 10. *In Situ* experimental system.

We can note that all measured reflection laws verify the specularity mode at low frequencies, but a diffuse reflexion behaviour is observed for high frequencies. The following figure shows the impulse response provided with a normal positionning of the microphone successively in front of the flat and the neoclassical surfaces. As the direct and specular peaks A and D are clearly readable as in the scale model, the different parts

of the facade response are more detailed in this measure, so we can discriminate the stonework (B), the windows (C) and the guardrail (E) backscattering contributions.

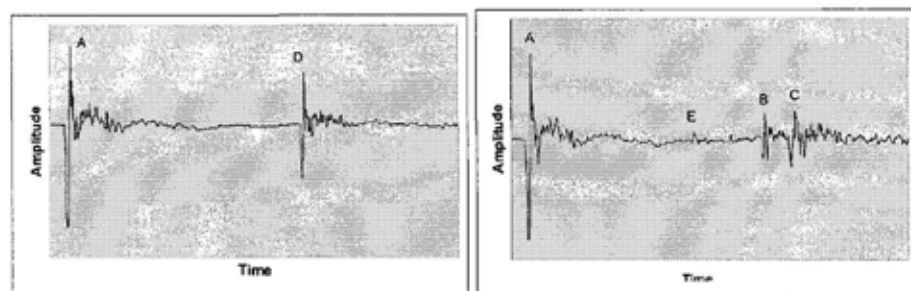


Figure 11. Comparative impulse responses from flat and neoclassical facades.

An accurate impulse response observation shows the correlation between the fractal analysis of the facades and the pure acoustical diffractive behaviour [15].

### 5.3 Results comparison

The following results represent the measured polar reflection laws for each specified frequency band. Values describe the early reflected sound energy level (the first 10 milliseconds backscattered signal) for each reception angle varying from  $15^\circ$  to  $165^\circ$ .

Correspondence between geometrical and metrological results is clearly visible through the « bilobe » behavior of the low frequencies, evolving towards a cardioïd figure for the medium acoustical domain, including three reflection lobes in the incidence direction and at angles  $30^\circ$ - $140^\circ$  (consequence from sound reflection on windows). As shown on figure 12, results for high frequencies confirm the acoustical validity of the fractal geometry evaluation model too, as we can read on both results the diffusivity peaks in the incidence direction and at characteristic angles from  $60^\circ$  to  $75^\circ$  ( $105^\circ$ - $170^\circ$ ), and for grazing angles from  $15^\circ$  to  $35^\circ$  ( $145^\circ$ - $165^\circ$ ).

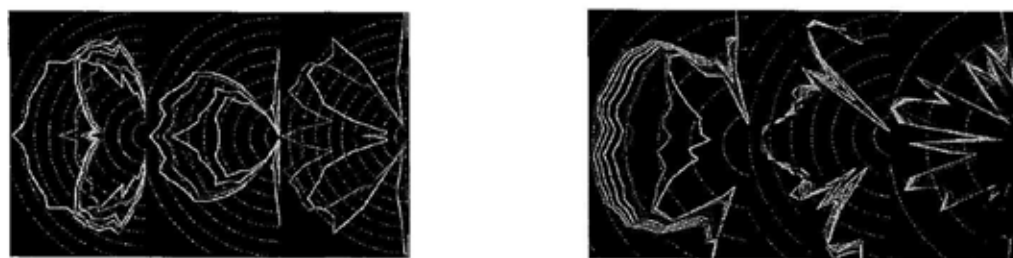


Figure 12. Comparison of fractal and measured result data around 100 Hz, 1 kHz and 4 kHz.

## 6 Conclusion

Through the determination of the structure factor, the Minkowski sausage technique provides a quantification of the scatter distribution function of the indented surface of a specified urban neo-classical facade at each incidence angle. In order to validate this model, experimental results have been compared to those indicators at each incidence angle, through an *in situ* MLS measurement. Comparison between results and



characterization shows a good agreement for frequencies from 15 Hz to 6 kHz at non-grazing incidence angles, for eight frequency bandwidths.

By discerning the angular vertex densitometry of main types of architectures, we will be able to compute their specific angular spatial scattering function for every frequency, directly from the Minkowski analysis of the numerical 3-D model of their geometry. Developed in the aim of architectural design tools for urban acoustics, this method allows a good evaluation of the acoustical reaction of an urban surface by using morphological attributes of an architecture. Through that morphological treatment of architectural shapes, this research work will confirm the definition of the diffusion process as a geometrical-dependant phenomenon, influenced by the built structure on urban acoustics.

## 7 References

1. Lord Rayleigh, The theory of Sound, Vol. 2, *Dover ed*, New-york, 1945, pp. 89-96.
2. Beckmann P., Spizzichino A., 1987 The Scattering of Electromagnetic Waves from Rough Surfaces, *Artec House, INC*, Norwood.
3. R. Makarewicz & P. Kokowski, Reflexion of noise from a building's facade, *Applied Acoustics* **43**, London, 1994, pp.149-157.
4. Lewers T, A Combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics, *Applied Acoustics* **38** (1993), pp. 161-178.
5. Hollander and al., Dynamic Structure Factor in a Random Diffusion Model, *Journal of Statistical Physics*, Vol.76, **5:6**, 1994.
6. Roman, 1997, Diffusion on self-similar structures, *Fractals*, World Scientific, Vol. 5, **3**, September 1997 pp.379-393.
7. Gouyet J.F. : Physique et structures fractales, Paris, *Masson ed*, 1992, 234 p.
8. Mandelbrot, B., Les objets fractals, Paris, *Flammarion ed*, 1992.
9. Roman H.E. : *Phys. Rev.* **E51**, 5422 (1995).
10. Woloszyn P. : Mesures multiéchelles du tissu urbain et paramétrage d'un modèle de diffusion acoustique en milieu construit, *ed. ENPC*, Marne-la-Vallée, 1997.
11. Young P., Parceval's Theorem, *Physics* **114 A**, 1999.
12. Pfeiffer P., Obert M. & Cole M.W. : Fractal BET and FHH theories of adsorption: a comparative study, *Proc. R. Soc. Lond.* **A423**, pp. 169-188, 1989.
13. Woloszyn P. : Squaring the circle : diffusion volume and acoustic behavior of a fractal structure, *Paradigms of Complexity*, M. M. Novak (ed.), World Scientific Publishing, Singapore, 2000, pp. 299-300.
14. Woloszyn P. Is Fractal Estimation of a Geometry worth for Acoustics ?, *Emergent Nature*, M. M. Novak (ed.), World Scientific Publishing Singapore, 2002, pp.423-425.
15. Woloszyn P., Suner B., Bachelier J. : Angular characterization of the urban facades diffusivity factor, *17<sup>th</sup> International Congress of Acoustics*, Rome, Italy, 2001.

# BINOMIAL MULTIPLICATIVE MODEL OF CRITICAL FRAGMENTATION

H. KATSURAGI, D. SUGINO, AND H. HONJO

*Department of Applied Science for Electronics and Materials  
Interdisciplinary Graduate School of Engineering Sciences  
Kyushu University, 6-1 Kasugakoen, Kasuga, Fukuoka 816-8580, Japan  
E-mail: katsurag@asem.kyushu-u.ac.jp*

We report the binomial multiplicative model for low impact energy fragmentation. Impact fragmentation experiments were performed for low impact energy region, and it was found that the weighted mean mass is scaled by the pseudo control parameter *multiplicity*. We revealed that the power of this scaling is a non-integer (fractal) value and has a multi-scaling property. This multi-scaling can be interpreted by a binomial multiplicative (simple biased cascade) model. Although the model cannot explain the power-law of fragment-mass cumulative distribution in fully fragmented states, it can produce the multi-scaling exponents that agree with experimental results well. Other models for fragmentation phenomena were also analyzed and compared with our model.

Keywords: brittle fragmentation, power-law, critical phenomena, multi-scaling

## 1 Introduction

Impact fragmentation of brittle solids is a typical nonlinear phenomenon. Small impact cannot make brittle solids cleave. However, large impact produces cracks irreversibly and makes brittle solids fissure to small pieces of fragments. This ubiquitous phenomenon can be seen even in our everyday lives. Thus, many scientists and engineers have studied this issue. As known well, cumulative distribution of fragment mass shows power-law<sup>1</sup>. Oddershede et al.<sup>2</sup> and Meibom and Balslev<sup>3</sup> investigated what controls the exponent of power-law distribution. They found that the exponent is determined by the dimensionality of fractured object. Ishii and Matsushita performed the 1-dimensional fragmentation experiments with long thin glass rods<sup>4</sup>. They dropped the glass rods from various heights. The cumulative distribution obeyed power-law form at high dropping height, and obeyed log-normal form at low one.

Recently, Kun and Herrmann investigated the damage-fragmentation transition for low impact energy collision by numerical simulation<sup>5</sup>. They used the granular solid disks colliding by a point to each other<sup>6</sup>. The transition from damaged state to fragmentation state was observed by increasing the relative collision speed. They measured the critical exponents of damage-fragmentation transition and realized that scaling-laws of the percolation universality are satisfied near this transition region. On the other hand, Åström et al. studied the low energy fragmentation using the random distorted lattice with elastic beam model and fluid MD model with LJ pair potential<sup>7</sup>. They corrected that the critical behavior for low energy fragmentation differs from that of percolation. Oddershede et al. said the fragmentation process is a kind of self-organized critical phenomenon<sup>2</sup>. However, most of experiments examined only on high imparted energy fragmentation. There are no

experiments on critical behavior of fragmentation by low imparted energy.

In order to study the critical fragmentation, we performed simple experiments of fragmentation. We considered a simple binomial multiplicative scenario of critical fragmentation. In this paper, we report on results of detailed analyses on the model. In the next section, experimental results are presented. In Sec. 3, we introduce a binomial multiplicative model and analyze it. In Sec. 4, the results are compared with other possible models.

## 2 Experiment

We used glass tube samples as 2-D fractured objects. The tube was put between a stainless stage and a stainless plate. A brass weight was dropped to the stainless plate. The falling height was controlled on slightly higher than the point at which samples did not fracture. After fragmentation, we collected fragments and measured the mass of each fragment with an electronic balance. Fractured tubes have the approximate 2-D geometry (50 mm outside diameter, 2 mm thick, and 50, 100, 150 mm length). More detailed experimental setups are described in Ref. 8.

According to Kun and Herrmann's result, the control parameter should be the imparted energy per unit sample mass  $\epsilon$ , and the order parameter should be the maximum fragment mass  $M_{\max}$ <sup>5</sup>. The  $\epsilon$  is calculated as  $\epsilon = M_w gh / M_{ob}$ , where  $M_w$ ,  $g$ ,  $h$ , and  $M_{ob}$  correspond to the mass of falling weight, the gravitational acceleration, the height of falling weight, and the mass of target sample, respectively. The log-log plot of maximum fragment mass  $M_{\max}$  vs. imparted energy per unit sample mass  $\epsilon$  is shown in Fig. 1. As can be seen in Fig. 1,  $M_{\max}$  and  $\epsilon$  relate with negative correlation, qualitatively. However, since the data in Fig. 1 contain large uncertainties, we cannot discuss quantitatively on the critical scaling by this plot. Therefore, we have to use another parameter to analyze quantitatively.

Campi proposed a pseudo control parameter *multiplicity*  $\mu$  in Ref. 9. The  $\mu$  is defined as,

$$\mu = m_{\min} \frac{M_0}{M_1}. \quad (1)$$

Where  $m_{\min}$ ,  $M_0$ , and  $M_1$  correspond to the smallest limit of fragment mass (we fix it at 0.01 g), the total number of fragments, and the total mass of the all fragments, respectively. The fragmentation critical point corresponds to the value  $\mu = 0$  by this definition. In general, we can introduce the  $k$ -th order moment of fragment mass distribution  $M_k$  as,

$$M_k = \sum_m m^k n(m), \quad (2)$$

where  $n(m)$  is the number of fragments of mass  $m$ . Certainly,  $M_0$  and  $M_1$  in Eq. (1) are specific cases of  $M_k$  ( $k = 0$  and  $1$ , respectively). We consider the  $k$ -th order weighted mean fragment mass  $M_{k+1}/M_k$ , and assume the scaling,

$$\frac{M_{k+1}}{M_k} \sim \mu^{-\sigma_k}. \quad (3)$$

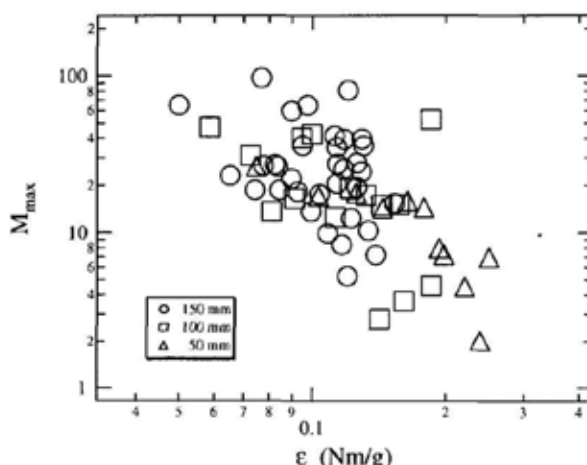


Figure 1. The maximum fragment mass  $M_{\max}$  vs. the imparted energy per unit sample mass  $\epsilon$ . Although the rough correlation between  $M_{\max}$  and  $\epsilon$  can be seen, it is too fluctuating to discuss quantitatively.

In Fig. 2, we show the log-log plot of  $M_2/M_1$  as a function of  $\mu$ . Contrary to the Fig. 1, the data in Fig. 2 seem to be fitted by a unified scaling line. The scaling result for  $k = 1$  is presented as a solid line in Fig. 2. We obtained the nontrivial scaling exponent  $\sigma_1 = 0.84 \pm 0.05$ . For other  $k$  regime, multi-scaling exponent values of  $\sigma_k$  were obtained as shown in Fig. 3 (circle marks). In spite of the trivial value  $\sigma_{k=0} = 1$ ,  $\sigma_k$  varies with  $k$ , and seems to approach to the nontrivial value ( $\approx 0.6$ ).

From the definition of  $\gamma_k$  as

$$\frac{M_k}{M_1} \sim \mu^{-k\gamma_k}, \quad (4)$$

the obtained  $\gamma_k$  values are plotted as square marks in Fig. 3. It seems that  $\gamma_k$  approaches to the value around 0.6 again. Of course, Eqs. (3) and (4) relate to each other. The relation  $\sum_{i=1}^{k-1} \sigma_i = k\gamma_k$  holds for any  $k$ . Thus, when the  $\sigma_k$  has a trivial value 1 for all  $k$ ,  $\gamma_k$  varies as  $(k-1)/k$ . The trivial curves are shown as broken lines in Fig. 3. In addition, the relation  $(k+1)\gamma_{k+1} - k\gamma_k = \sigma_k$  can be computed from Eqs. (3) and (4). If the difference between  $\gamma_{k+1}$  and  $\gamma_k$  becomes small,  $\gamma_k$  and  $\sigma_k$  approximately have the same value, as seen in Fig. 3 for large  $k$ .

On the other hand, when the imparted energy was sufficient to fully shatter, many fragments were created and the cumulative distribution of fragment mass obeyed power-law form. Our results suggest the power is 0.5 about the 2-D fragmentation with the flat impact<sup>8</sup>. In this regime, cumulative distribution functions are collapsed by the scaling function written as  $N(m)/M_0 \sim f(m/\mu^{-\sigma})$ . The function  $f(x)$  consists of the scaling part  $f(x) \sim x^{-0.5}$  and the decaying part due to the

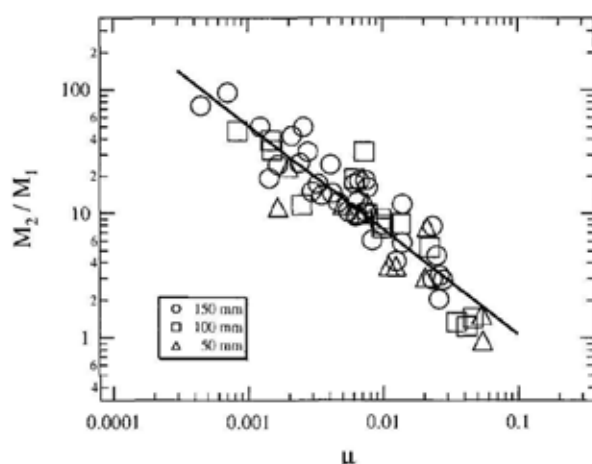


Figure 2. The weighted mean fragment mass  $M_2/M_1$  vs. the multiplicity  $\mu$ . The solid line indicates the form of the power-law fitting  $M_2/M_1 \sim \mu^{-\sigma_1}$  ( $\sigma_1 = 0.84 \pm 0.05$ ). Three different size results are fitted by the unified scaling independently of size.

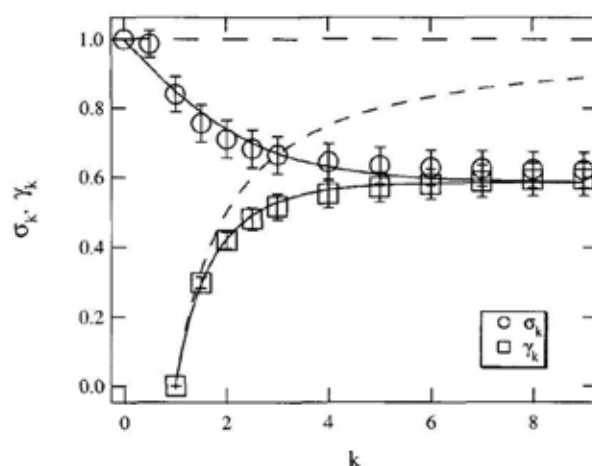


Figure 3. Multi-scaling exponent  $\sigma_k$  and  $\gamma_k$  obtained from  $k$ -th order weighted mean fragment mass. The broken lines indicate the case of trivial integral scaling corresponding to  $\sigma_k = 1$  and  $\gamma_k = (k-1)/k$ . The solid lines depict results of the binomial multiplicative model at  $a = 2/3$ . The  $\sigma_k$  and  $\gamma_k$  asymptotically approach to the same value, in large  $k$  range. From the definition of Eqs. (3) and (4), the values  $\sigma_{k=0}$  and  $\gamma_{k=1}$  exactly show 1 and 0, respectively.

finite size effect. The value 0.5 concurs to the results of Hayakawa<sup>10</sup> and Åström et al.<sup>7</sup>. In contrast, this value is not consistent with the percolation scaling ansatz<sup>11</sup>. In the percolation scaling ansatz, the scaling exponent of cluster size cumulative distribution must be greater than 1. Therefore, we can consider that the universality classes of critical fragmentation and percolation criticality are different each other.

### 3 Model

In order to explain this multi-scaling property, we introduce a simple biased cascade model. A binomial multiplicative process is considered with a unit mass initial condition. Here we consider a asymmetrical cleaving presented by a parameter  $a$ . We can limit the range of the parameter  $a$  as  $1/2 \leq a \leq 1$  by the symmetry of the model. The initial unit mass is divided into two fragments whose masses are  $a$  and  $1 - a$  at first step. This biased cleaving continues some steps until the imparted energy dissipates. In this model, we can easily calculate the exponents  $\sigma_k$  and  $\gamma_k$  from Eqs. (3) and (4) as

$$\frac{a^{k+1} + (1-a)^{k+1}}{a^k + (1-a)^k} = 2^{-\sigma_k}, \quad (5a)$$

$$a^k + (1-a)^k = 2^{-k\gamma_k}, \quad (5b)$$

or more explicit forms as,

$$\sigma_k = -\frac{\ln[a^{k+1} + (1-a)^{k+1}] - \ln[a^k + (1-a)^k]}{\ln 2}, \quad (6a)$$

$$\gamma_k = -\frac{\ln[a^k + (1-a)^k]}{\ln 2}. \quad (6b)$$

If we choose a value  $a = 2/3$ , the  $\sigma_k$  and  $\gamma_k$  become the values depicted by the solid lines in Fig. 3. One can confirm the pretty good agreement with experimental data. The trivial case presented by broken lines in Fig. 3 corresponds to the case  $a = 1/2$ . In this case, all fragments at each step are perfectly equal. In the case  $a \neq 1/2$ , the fragment size distribution has variation and exhibits multifractal scaling.

This model is so simple that we can calculate the fragment mass and the number of fragments exactly. We consider the  $s$ -th step, and suppose the fragments in which the factor  $a$  (or  $1 - a$ ) works  $t$  (or  $s - t$ ) times. In such fragments, the mass  $m_s(t)$  is written as,

$$m_s(t) = a^t(1-a)^{s-t}, \quad \left(\frac{1}{2} \leq a \leq 1\right). \quad (7)$$

And the number of fragments  $n_s(t)$  is described as,

$$n_s(t) = \frac{s!}{t!(s-t)!}. \quad (8)$$

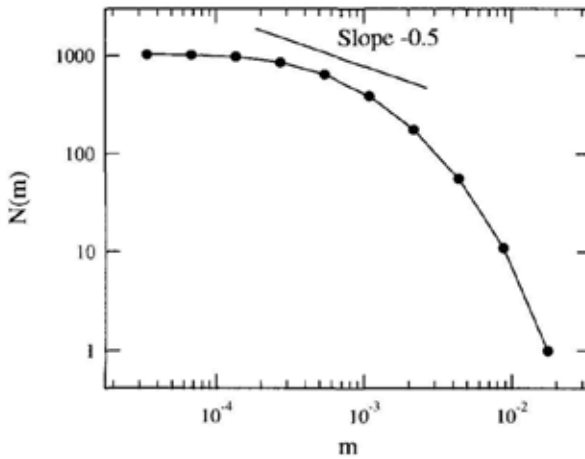


Figure 4. The cumulative fragment mass distribution for the binomial multiplicative model with  $a = 2/3$  and  $s = 10$ .

Since we are interested in the cumulative distribution of fragment mass, the cumulative number of fragments  $N_s(t)$  is calculated as

$$N_s(t) = \int_t^\infty n_s(t') dt' = \sum_{i=t}^s \frac{s!}{i!(s-i)!}. \quad (9)$$

In Fig. 4, we show the cumulative distribution of fragment mass obtained from the model. The parameters are taken as  $a = 2/3$  and  $s = 10$ . The line of slope  $-0.5$  corresponding to the experimental result is also shown as a solid line in Fig. 4. Unfortunately, clear power-law, which follows the experimental data, cannot be observed. However, the distribution curve in Fig. 4 seems to include the region of slope  $-0.5$ . In this model, we can calculate the local power  $\tau - 1$  directly by the relation,

$$\frac{N_s(t-1)}{N_s(t)} = \left[ \frac{m_s(t-1)}{m_s(t)} \right]^{-(\tau-1)}. \quad (10)$$

Solving the Eq. (10), we obtain the exact form of  $\tau - 1$  as follows,

$$\tau - 1 = - \frac{\ln \left[ \frac{N_s(t-1)}{N_s(t)} \right]}{\ln \left[ \frac{m_s(t-1)}{m_s(t)} \right]} = - \frac{\ln \left[ \frac{\sum_{i=t-1}^s \frac{s!}{i!(s-i)!}}{\sum_{i=t}^s \frac{s!}{i!(s-i)!}} \right]}{\ln \left[ \frac{a^{t-1}(1-a)^{s-t+1}}{a^t(1-a)^{s-t}} \right]}. \quad (11)$$

We show the relations among  $\tau - 1$ ,  $s$ , and  $t$  in Fig. 5. As can be seen in Fig. 5, the lower limit of the local slope  $\tau - 1$  is 0, and it has a divergent tendency. The value 0.5 is not a particular one.

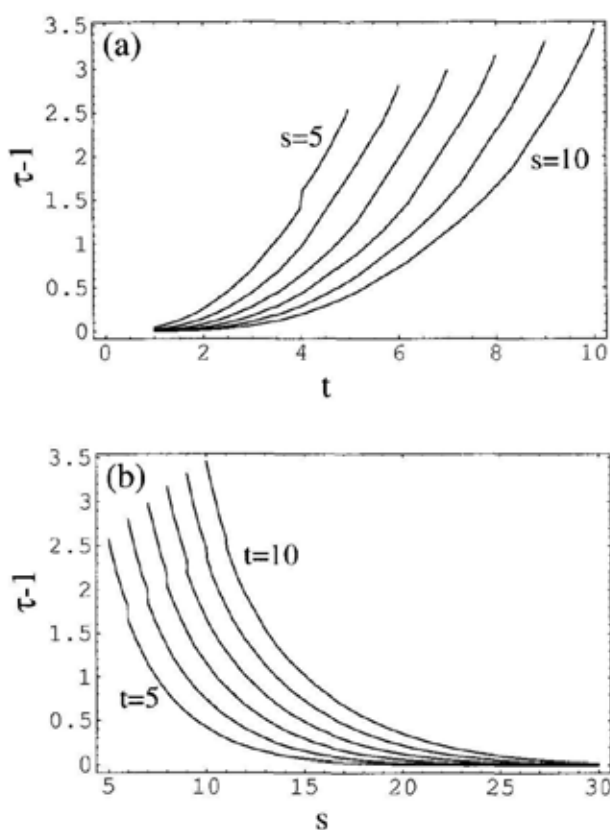


Figure 5. The local scaling exponent of cumulative fragment mass distribution  $\tau - 1$ . (a) The relations between  $\tau - 1$  and  $t$  are shown. Each curve corresponds to the case  $s = 5, 6, \dots, 10$  from left to right. (b) The relations between  $\tau - 1$  and  $s$  are shown. Each curve corresponds to the case  $t = 5, 6, \dots, 10$  from left to right.

#### 4 Discussion

In Sec. 2, we concluded that the universality of the critical fragmentation differs from that of percolation. Instead, the weighted mean fragment mass was studied to reveal the universality of the critical fragmentation. It indicates the multi-scaling nature and is modeled by the simple binomial multiplicative model. There are some other candidates for the critical fragmentation. From now on, we discuss and compare them.

Similar multiplicative model for turbulent flows was proposed by Meneveau and Sreenivasan<sup>12</sup>. They investigated the energy cascade of eddies, and obtained good agreement with experimental data at  $a = 0.3$  (in their paper, the corresponding parameter was written as  $p_1$ ). This value slightly coincides to ours  $1 - a = 1/3$ .



The same physical mechanism might dominate both cascades of fragmentation and turbulence.

We can fit the data by  $a = 2/3$  very well indeed, however, the reason of symmetry breaking by  $a \neq 1/2$  is not understood well. While the model always requires the exact asymmetry presented by  $a$ , the cleaving point might distribute. We can consider the simple distributed model as described below. We set the unit mass segment initial condition again, and consider the probability  $p(x)dx$  which presents the cleaving point in the range from  $x$  to  $x + dx$  at each step. We assume the symmetrical distribution as  $p(x) = 4x$  ( $0 \leq x \leq 1/2$ ), and  $4 - 4x$  ( $1/2 \leq x \leq 1$ ). This is one of the simplest distribution presented by isosceles triangles. The normalization condition of this model is  $\int_0^1 p(x)dx = 1$ . In this model, we can calculate the expectation value of the cleaving point  $x$  (or  $1 - x$ ) as,

$$\int_0^{1/2} (1-x)4xdx + \int_{1/2}^1 x(4-4x)dx = \frac{2}{3}. \quad (12)$$

Note that we cannot distinguish the cleaving state  $(x, 1-x)$  and the state  $(1-x, x)$ . Thus, the  $x$  can be limited in the region  $1/2 \leq x \leq 1$ . The expectation value is nearly the same as one ( $a = 2/3$ ) of the above mentioned multiplicative model. We can also calculate the  $k$ -th order moment  $M_k$  as,

$$M_k = \int_0^1 [x^k + (1-x)^k]p(x)dx = \frac{8}{(k+1)(k+2)} \left[ 1 - \left(\frac{1}{2}\right)^{k+1} \right]. \quad (13)$$

We show the  $\sigma_k$  computed from the Eqs. (13), (3), and (1) as a solid line in Fig. 6. The result does not supply the agreement with the experimental data, particularly in large  $k$  range.

Matsuhita<sup>13</sup> and Turcotte<sup>14</sup> introduced the model for power-law fragmentation. Matsushita examined the model in which each fragment cleaves into 4 pieces at each step. Then 1 piece does not break any more, and the other 3 pieces cleave into 4 sub-pieces at next step. The same procedure works upon all sub-pieces at each step. According to his model, the exponent of the power-law cumulative distribution of fragment mass becomes  $\tau - 1 = \ln 3 / \ln 4 \simeq 0.79$ . We can easily modify this model to the case  $\tau - 1 = \ln 2 / \ln 4 = 1/2$  by changing the remaining piece number 1 into 2. In this condition, we can calculate the  $M_k$  for this modified version of the partial remaining model at  $s$ -th step as,

$$M_k = \sum_{i=1}^s 2^i \left(\frac{1}{4}\right)^{ik}. \quad (14)$$

Then, the  $\sigma_k$  can be computed again, however, the value of  $\sigma_k$  depends not only on  $k$ , but also on  $s$ . We show the  $\sigma_k$  obtained from this model at  $s = 10$  as a broken line in Fig. 6. This model also cannot provide the appropriate curve of the  $\sigma_k$ . Thus the exact  $a = 2/3$  binomial multiplicative model seems to be the most possible model in terms of multi-scaling exponents  $\sigma_k$ .

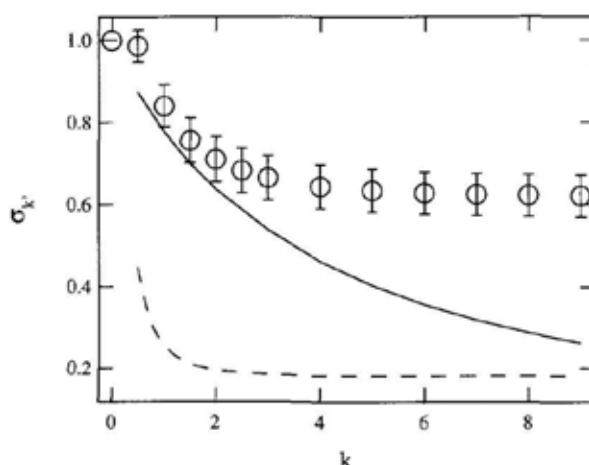


Figure 6. Comparison between the experimental data and the other considerable models. The experimental data are shown as circle marks. The exponent  $\sigma_k$  obtained from the symmetric (isosceles triangle) distribution model is presented by the solid line. The broken line indicates the result of the modified partial remaining model at  $s = 10$ .

## 5 Conclusions

We investigated the criticality of brittle fragmentation. Some models to interpret the experimental result are proposed. The exact binomial multiplicative model can produce the adequate approximation for the exponent  $\sigma_k$ . And the cumulative distribution obtained from the model is not so worse. However, it requires that the cleaving point is exactly at  $a$ . Since the isosceles triangle model has non-divergent standard deviation, the distribution of fragment mass resulting from the model must approach to the log-normal distribution due to the central limit theorem<sup>15,16</sup>. In addition, its  $\sigma_k$  differs from the experimental data, particularly in large  $k$  region. The modified partial remaining model can explain the experimental value of the exponent  $\tau$  very well. However, the  $\sigma_k$  from the model shows large discrepancy from the experimental data. Each model has merits and demerits. The totally sufficient model is not presented yet. Furthermore, these scaling exponents will depend on detail load condition and dimensionality of fractured object. More detailed experiments and analyses are necessary to solve the criticality of brittle fragmentations completely.

## Acknowledgments

The authors would like to thank Professor J. Timonen, Dr. J. A. Åström, Professor S. Ohta, Professor H. Sakaguchi, and Professor K. Nomura for useful discussion and comments.

## References

1. D. Beysens, X. Campi, and E. Pefferkorn eds., *Fragmentation Phenomena* (World Scientific, Singapore, 1995).
2. L. Oddershede, P. Dimon, and J. Bohr, *Phys. Rev. Lett.* **71**, 3107 (1993).
3. A. Meibom and I. Balslev, *Phys. Rev. Lett.* **76**, 2492 (1996).
4. T. Ishii and M. Matsushita, *J. Phys. Soc. Jpn.* **61**, 3474 (1992).
5. F. Kun and H. J. Herrmann, *Phys. Rev. E* **59**, 2623 (1999).
6. F. Kun and H. J. Herrmann, *Int. J. Mod. Phys. C* **7**, 837 (1996).
7. J. A. Åström, B. L. Holian, and J. Timonen, *Phys. Rev. Lett.* **84**, 3061 (2000).
8. H. Katsuragi, D. Sugino, and H. Honjo, *Phys. Rev. E* **68**, 046105 (2003), preprint [cond-mat/0307756].
9. X. Campi, *Phys. Lett. B* **208**, 351 (1988); X. Campi and H. Krivine, *Observables in Fragmentation*, pp.312 in Ref. 1.
10. Y. Hayakawa, *Phys. Rev. B* **53**, 14828 (1996).
11. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, London, 1992).
12. C. Meneveau and K. R. Sreenivasan, *Phys. Rev. Lett.* **59**, 1424 (1987).
13. M. Matsushita, *J. Phys. Soc. Jpn.* **54**, 857 (1985).
14. D. L. Turcotte, *J. GeoPhys. Res.* **91**, 1921 (1986).
15. M. Matsushita and K. Sumida, *Bull. Facul. Sci. & Eng. Chuo Univ.* **31**, 69 (1988).
16. A. Diehl, H. A. Carmona, L. E. Araripe, J. S. Andrade, Jr., and G. A. Farias, *Phys. Rev. E*, **62**, 4742 (2000); J. A. Åström, R. P. Linna, and J. Timonen, *ibid.*, **65**, 048101 (2002); A. Diehl, S. Andrade, Jr., and G. A. Farias, *ibid.*, **65**, 048102 (2002).

# STUDY ON THE IMPROVED FRACTAL INTERPOLATION SURFACE OF THE ATTITUDE AND SURFACE OF FAULT

HONGQUAN SUN

*School of Civil Engineering, Hebei University of Technology, Tianjin 300130, P. R. China*

*E-mail: hqsun@public.tpt.tj.cn*

HEPING XIE

*Institute of Rock Mechanics and Fractals, China University of Mining and Technology, Beijing  
100083, P. R. China*

*E-mail: xiehp@mail.cumt.edu.cn*

This paper consists of three parts. The first part describes a mathematical model of fractal interpolation surface on a rectangle field and the calculation formula of the fractal interpolation surfaces. The second part presents the study of attitude (the trend and the obliquity) of the fault by using multivariate statistics. The third part discusses the simulation of roughness of the fault surface dealing with the improved methods, the partition of local domains and the determination of vertical scaling factor, of fractal interpolation surfaces. At the same time, the fractal dimension of the interpolated fault surface is obtained. The theory and method discussed in this paper provides a new way for studying the influence of the roughness of the fault surface in mining engineering and civil engineering.

## 1. Introduction

The accidents of slope instability and roof fall in mining engineering and civil engineering occur frequently. The occurrence of the accidents is closely related to the influence of faults and joints in rocks. The attitude and the surface roughness of the faults and the joints affect the occurrences of the slope stability and the roof cave-in directly. For many years, the researchers in the field of rock mechanics, geology and mining have been paying much attention to the study of the shape and the surface roughness of the faults and joints.

However, faults and joints are in the different layers underground and it is difficult to obtain the shape and roughness of the fault surfaces. So, it is urgent to develop mathematical models (Mandelbrot 1982, Barnsley 1986, 1988, Massopust 1994, Falconer 1990) so that the real fault shapes can be interpolated approximately for analyzing, simulating and predicting in order to research the influences of the roughness of faults and joints on the accidents in mining engineering and civil engineering.

Professor H. Xie(1998) has found out the fractal properties of the fault surface and pointed out that the fault surfaces have fine fractal features of statistical self-similarity and the rough shape of the fault surfaces is related to the lithology and tectonic stress characteristics. He concluded that the roughness of fault surface influences the mining subsidence and the instability in the civil engineering directly.

In this paper, based on the principles of multivariate statistics and the theories and methods of fractal interpolation surfaces, the attitude and the roughness shape of fault surface are simulated and the fractal dimension of the simulated fault surface is obtained by using the altitude data of the fault surface.

## 2. The principles of fractal interpolation surface on a rectangle field

Let  $I = [a, b]$ ,  $J = [c, d]$ ; and  $D = I \times J = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$ . Subdivide  $D$  into the subintervals:

$$\begin{cases} a = x_0 < x_1 < \cdots < x_N = b \\ c = y_0 < y_1 < \cdots < y_M = d \end{cases} \quad (1)$$

Given a set of data on the grid:  $\{(x_i, y_j, z_{i,j}), i = 0, 1, \dots, N, j = 0, 1, \dots, M\}$ , we construct an interpolation function  $f: D \rightarrow R$ , such that  $f(x_i, y_j) = z_{i,j}$ , ( $i = 0, 1, \dots, N, j = 0, 1, \dots, M$ ).

We will discuss on the three dimensional domain (Heping Xie and Hongquan Sun 1997)  $K = D \times [h_1, h_2]$  ( $-\infty < h_1 < h_2 < +\infty$ ). For  $(c_1, d_1, e_1), (c_2, d_2, e_2) \in K$ , let  $d((c_1, d_1, e_1), (c_2, d_2, e_2)) = \max\{|c_1 - c_2|, |d_1 - d_2|, |e_1 - e_2|\}$ .

Let  $I_n = [x_{n-1}, x_n], J_m = [y_{m-1}, y_m], D_{n,m} = I_n \times J_m, n \in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\}$ . And let  $\Phi_n: I \rightarrow I_n, \Psi_m: J \rightarrow J_m$  be contraction mapping and satisfy:

$$\begin{cases} \Phi_n(x_0) = x_{n-1}, \quad \Phi_n(x_N) = x_n \\ \Psi_m(y_0) = y_{m-1}, \quad \Psi_m(y_M) = y_m \\ \left| \Phi_n(c_1) - \Phi_n(c_2) \right| < k_1 |c_1 - c_2| \\ \left| \Psi_m(d_1) - \Psi_m(d_2) \right| < k_2 |d_1 - d_2| \end{cases} \quad (2)$$

where  $c_1, c_2 \in I, d_1, d_2 \in J, 0 \leq k_1 < 1, 0 \leq k_2 < 1$ .

Let  $L_{n,m}: D \rightarrow R^2$  be a contraction transformation:  $L_{n,m}(x, y) = (\Phi_n(x), \Psi_m(y))$ . Let  $F_{n,m}: K \rightarrow [h_1, h_2]$  be continuous, which must obey four equations:

$$\begin{cases} F_{n,m}(x_0, y_0, z_{0,0}) = z_{n-1,m-1} \\ F_{n,m}(x_N, y_0, z_{N,0}) = z_{n,m-1} \\ F_{n,m}(x_0, y_M, z_{0,M}) = z_{n-1,m} \\ F_{n,m}(x_N, y_M, z_{N,M}) = z_{n,m} \end{cases} \quad (3)$$

For any  $(x_1, y_1), (x_2, y_2) \in D$  and  $z_1, z_2 \in [h_1, h_2]$ , we have

$$\begin{aligned} |F_{n,m}(x_1, y_1, z_1) - F_{n,m}(x_2, y_2, z_2)| &\leq k_3 |z_1 - z_2| \\ n \in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\}, 0 &\leq k_3 < 1 \end{aligned} \quad (4)$$

Let  $\Phi_n(x) = a_n x + b_n$ . With the conditions (2), we have

$$a_n x_0 + b_n = x_{n-1}, \quad a_n x_N + b_n = x_n \quad (5)$$

and obtain

$$\begin{cases} a_n = (x_n - x_{n-1}) / (x_N - x_0) \\ b_n = (x_{n-1} x_N - x_n x_0) / (x_N - x_0) \end{cases} \quad (6)$$

$$\therefore \quad \Phi_n(x) = x_{n-1} + \frac{x_n - x_{n-1}}{x_N - x_0}(x - x_0) \quad n \in \{1, 2, \dots, N\} \quad (7)$$

Let  $\Psi_m(y) = c_my + d_m$ . Similarly, with the conditions (2), we have

$$\begin{cases} c_m = (y_m - y_{m-1}) / (y_M - y_0) \\ d_m = (y_{m-1}y_M - y_my_0) / (y_M - y_0) \end{cases} \quad (8)$$

$$\therefore \quad \Psi_m(y) = y_{m-1} + \frac{y_m - y_{m-1}}{y_M - y_0}(y - y_0) \quad m \in \{1, 2, \dots, M\} \quad (9)$$

Let

$$\begin{aligned} F_{n,m}(x, y, z) &= e_{n,m}x + f_{n,m}y + g_{n,m}xy + s_{n,m}z + k_{n,m} \\ n &\in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\} \end{aligned} \quad (10)$$

According to Eqs. (3), we have

$$\begin{cases} z_{n-1,m-1} = e_{n,m}x_0 + f_{n,m}y_0 + g_{n,m}x_0y_0 + s_{n,m}z_{0,0} + k_{n,m} \\ z_{n,m-1} = e_{n,m}x_N + f_{n,m}y_0 + g_{n,m}x_Ny_0 + s_{n,m}z_{N,0} + k_{n,m} \\ z_{n-1,m} = e_{n,m}x_0 + f_{n,m}y_M + g_{n,m}x_0y_M + s_{n,m}z_{0,M} + k_{n,m} \\ z_{n,m} = e_{n,m}x_N + f_{n,m}y_M + g_{n,m}x_Ny_M + s_{n,m}z_{N,M} + k_{n,m} \end{cases} \quad (11)$$

Let  $s_{n,m}$  ( $n \in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\}$ ) be any real number and satisfy  $|s_{n,m}| < 1$  which is called a vertical scaling factor. We find that we can always solve the above equations for  $e_{n,m}, f_{n,m}, g_{n,m}$  and  $k_{n,m}$  in terms of the interpolation data and  $s_{n,m}$ . We obtain

$$\begin{cases} g_{n,m} = \frac{z_{n-1,m-1} - z_{n-1,m} - z_{n,m-1} + z_{n,m} - s_{n,m}(z_{0,0} - z_{N,0} - z_{0,M} + z_{N,M})}{x_0y_0 - x_Ny_0 - x_0y_M + x_Ny_M} \\ e_{n,m} = \frac{z_{n-1,m-1} - z_{n,m-1} - s_{n,m}(z_{0,0} - z_{N,0}) - g_{n,m}(x_0y_0 - x_Ny_0)}{x_0 - x_N} \\ f_{n,m} = \frac{z_{n-1,m-1} - z_{n-1,m} - s_{n,m}(z_{0,0} - z_{0,M}) - g_{n,m}(x_0y_0 - x_0y_M)}{y_0 - y_M} \\ k_{n,m} = z_{n,m} - e_{n,m}x_N - f_{n,m}y_M - s_{n,m}z_{N,M} - g_{n,m}x_Ny_M \end{cases} \quad (12)$$

$$n \in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\}$$

We define a new mapping  $G_{n,m}(x, y, z)$ :

$$G_{n,m}(x,y,z) = \begin{cases} F_{n,m}(x,y,z) + \alpha_{n,m}(y,z) \frac{x-x_{N-1}}{x_N-x_{N-1}}, & \begin{matrix} x \in [x_{N-1}, x_N] \\ n \in \{1, 2, \dots, N-1\} \\ m \in \{1, 2, \dots, M\} \end{matrix} \\ F_{n+1,m}(x,y,z) - \alpha_{n,m}(y,z) \frac{x_1-x}{x_1-x_0}, & \begin{matrix} x \in [x_0, x_1] \\ n \in \{1, 2, \dots, N-1\} \\ m \in \{1, 2, \dots, M\} \end{matrix} \\ F_{n,m}(x,y,z) + \beta_{n,m}(x,z) \frac{y-y_{M-1}}{y_M-y_{M-1}}, & \begin{matrix} y \in [y_{M-1}, y_M] \\ n \in \{1, 2, \dots, N\} \\ m \in \{1, 2, \dots, M-1\} \end{matrix} \\ F_{n,m+1}(x,y,z) - \beta_{n,m}(x,z) \frac{y_1-y}{y_1-y_0}, & \begin{matrix} y \in [y_0, y_1] \\ n \in \{1, 2, \dots, N\} \\ m \in \{1, 2, \dots, M-1\} \end{matrix} \\ F_{n,m}(x,y,z), & \text{others} \end{cases} \quad (13)$$

where

$$\alpha_{n,m}(y,z) = (F_{n+1,m}(x_0,y,z) - F_{n,m}(x_N,y,z)) / 2 \quad (14)$$

$$\beta_{n,m}(x,z) = (F_{n,m+1}(x,y_0,z) - F_{n,m}(x,y_M,z)) / 2 \quad (15)$$

According to the Eq. (13), we can define an Iterate Function System (IFS) (Barnsley 1986, 1988)  $W_{n,m}(x,y,z)$  on the field  $K$ :

$$W_{n,m}(x,y,z) = (\Phi_n(x), \Psi_m(y), G_{n,m}(x,y,z)), \\ n \in \{1, 2, \dots, N\}, m \in \{1, 2, \dots, M\}. \quad (16)$$

For such defined IFS, we have a unique attractor  $G = \{(x,y,f(x,y)) : (x,y) \in D\}$  which is the graph of a continuous function  $f$ , such that

$$f(x_i, y_j) = z_{i,j}; \quad i = 0, 1, \dots, N, \quad j = 0, 1, \dots, M. \quad (17)$$

Based on the Iterate Function System (16), the function of the self-affine fractal interpolation surface can be obtained:

$$f(x,y) = e_{n,m} \Phi_n^{-1}(x) + f_{n,m} \Psi_m^{-1}(y) + g_{n,m} \Phi_n^{-1}(x) \Psi_m^{-1}(y) \\ + s_{n,m} f(\Phi_n^{-1}(x), \Psi_m^{-1}(y)) + k_{n,m} \quad (18)$$

where  $e_{n,m}$ ,  $f_{n,m}$ ,  $g_{n,m}$  and  $k_{n,m}$  are obtained from Eqs.(12) and

$$\Phi_n^{-1}(x) = (x - x_{n-1}) \frac{x_N - x_0}{x_n - x_{n-1}} + x_0, \quad x \in [x_{n-1}, x_n], \quad n \in \{1, 2, \dots, N\}, \quad (19)$$

$$\Psi_m^{-1}(y) = (y - y_{m-1}) \frac{y_M - y_0}{y_m - y_{m-1}} + y_0, \quad y \in [y_{m-1}, y_m], \quad m \in \{1, 2, \dots, M\}, \quad (20)$$

Let  $N = M$  in the Eq. (18) and  $a_n = c_m = 1/N$  in Eqs. (6) and (8). So that :

$$\begin{cases} \Phi_n = \frac{1}{N} x + b_n, & n \in \{1, 2, \dots, N\} \\ \Psi_m = \frac{1}{N} y + d_m, & m \in \{1, 2, \dots, N\} \end{cases} \quad (21)$$

The number of the fields for the interpolation is  $N^2 (= N \times N)$ . Suppose  $f^*$  is the fractal interpolation function, then we have the dimension theorem of fractal interpolation function as follows:

Suppose  $\sum \Sigma |s_{n,m}| > N$  and the interpolation points are noncoplanar, then the box dimension of fractal interpolation surface is given by :

$$\dim(\text{graph } f^*) = 1 + \log_N \sum_{n=1}^N \sum_{m=1}^N |s_{n,m}|. \quad (22)$$

### 3. Attitude analyses of the fault surface

#### 3.1 Elevation data of the fault surface

We obtained 28 elevation data of a fault surface with seismic reflection data in a coal field in south China. For simplicity, we move the down-left corner of the research field to the origin of the coordinate system.

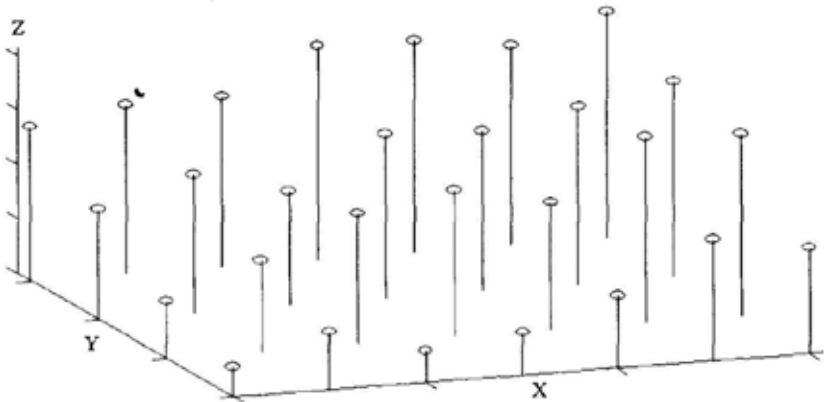


Figure 1. Spatial distribution of the elevation data on the fault surface



Figure 1 gives the spatial distribution of the 28 point elevation data. The roundlets show the locations of the data on the fault surface. The mesh shape of the data is shown in Figure 2.



Figure 2. Mesh of the elevation data on the fault surface

On the statistics, the elevation data of fault surface include two kinds of information: the attitude and the roughness of the fault surface. At first we divide the elevation data of the fault surface into the attitude and the roughness of the fault. We study the attitude of the fault with multivariate statistics analyses and simulate the roughness of the fault with the fractal interpolation surface. Now we put up the attitude analyses of the fault surface.

### 3.2 Trend of the fault

The study of the attitude of the fault is based on the principle of the trend surface analyses. We use the practical data obtained from the fault surfaces to fit the first order trend surface.

$$\hat{z} = b_0 + b_1x + b_2y \quad (23)$$

Where  $x$  and  $y$  are coordinates,  $\hat{z}$  trend value of the fault surface, and  $b_0$ ,  $b_1$  and  $b_2$  coefficients of the trend surface, which can be obtained by using the least square method.

The intersect line of trend surface (Eq.(23)) and the plane  $z = 0$  is the direction of fault surface. The angle formed by the first order trend surface and the plane  $XY$  is the obliquity of the fault surface.

On the practical data on the fault surface, we obtained the first trend surface equation:

$$\hat{z} = 556.7590 + 0.5454x + 1.4678y \quad (24)$$

Figure 3 shows the shape of the trend surface.

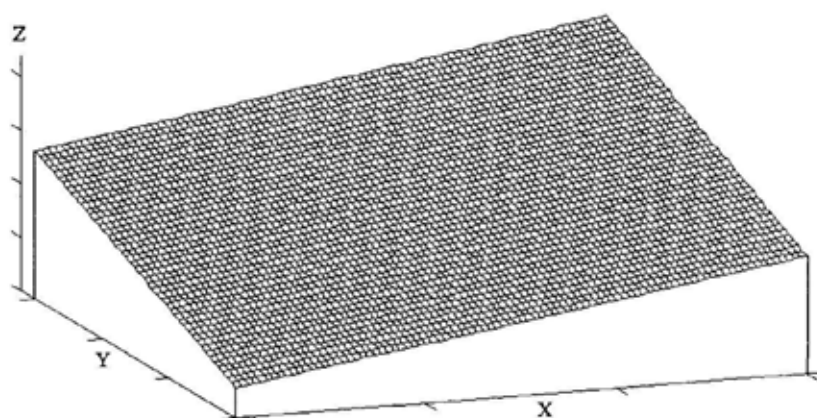


Figure 3. The first order trend surface of the fault surface

According to the analyses of the trend of the fault, let the  $z$  value of the first order trend surface Eq. (24) be zero. The equation of trend line can be obtained (see Fig. 4)

$$y = -0.3716x - 379.3153 \quad (25)$$

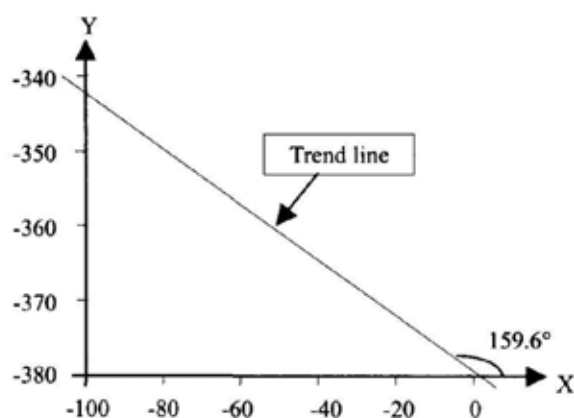


Figure 4. Trend line of the fault surface

Suppose the angle formed by trend line and  $X$  axis is  $\theta$ , then :

$$\theta = 180^\circ - \tan^{-1}(0.3716) = 180^\circ - 20.4^\circ = 159.6^\circ \quad (26)$$

Let the angle between trend line and  $Y$  axis be  $\alpha$ , then:

$$\alpha = \theta - 90^\circ = 69.6^\circ \quad (27)$$

So, the trend of the fault is NWW  $69.6^\circ$ . Figure 4 shows the trend line of the fault on the  $XY$  plan.

### 3.3 Obliquity of the fault

The angle formed by the first order trend surface and XY plane expresses the obliquity of the fault surface. Based on the calculating way of the angle of the two planes, we can obtain the angle of the first trend surface and XY plane.

Let  $P_1$  be the first order trend surface and  $P_2$  be XY plane, that is  $P_1: z = 556.75090 + 0.5454x + 1.4678y$ ,  $P_2: z = 0$ , then the angle between  $P_1$  and  $P_2$ ,  $\phi$ , can be expressed as

$$\cos \phi = \frac{1}{\sqrt{0.5454^2 + 1.4678^2 + 1}} = \frac{1}{1.8579} = 0.5283 \quad (28)$$

New we have:

$$\phi = 57.4^\circ \quad (29)$$

So, we obtained that the obliquity of the fault surface is  $57.4^\circ$ .

## 4. Improved fractal interpolation surface of the fault surface

The surface interpolated directly by the equations of fractal interpolation surface gives the characteristics of the strict self-similarity fractals. In fact, objects studied in nature seldom exhibit strict self-similarity fractals. Fractals in nature have the statistical fractal feature usually. Similarly, fault surface has the fractal feature but not strict self-similarity. So we put forward the methods of the improved self-affine fractal interpolation with the ways of the partition of local field and selecting of vertical scaling factor.

### 4.1 The partition of local field

The fault surface possesses fractal character, but not strict self-similarity. Researches indicate that the variable  $z(x)$  reflecting roughness of the fault surface is a regional variable, that is it includes both pertinence and randomness. Using the variogram theory of geostatistics (Hongquan Sun 1990), we put forward the method of the local field partition. The expression of the spherical model of variogram is:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ c_0 + c \left( \frac{3}{2} \cdot \frac{h}{a} - \frac{1}{2} \cdot \left( \frac{h}{a} \right)^3 \right) & 0 < h \leq a \\ c_0 + c & h > a \end{cases} \quad (30)$$

where  $a$  is range,  $c_0$  nugget and  $c_0 + c$  sill (see Figure 5).

The physical meaning of range  $a$  is that if the distance between two points is less than  $a$ , the variation of this two points is related to the distance between them. In the process of fractal interpolation, the range  $a$  is used as the basis of the partition of the local field. The calculation formulas of the fractal interpolation are used in the local fields.

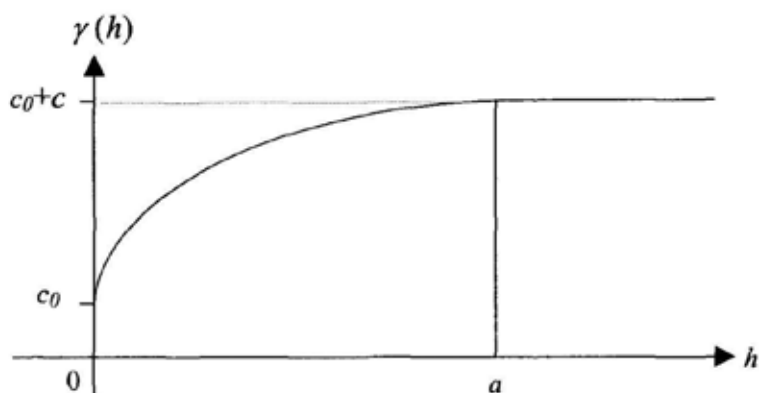


Figure 5. Spherical variogram model

Suppose we found out  $a = 0.7$  of a set of data with the interval 0.22, then we select  $4 \times 4$  points as the local field (see Figure 6).

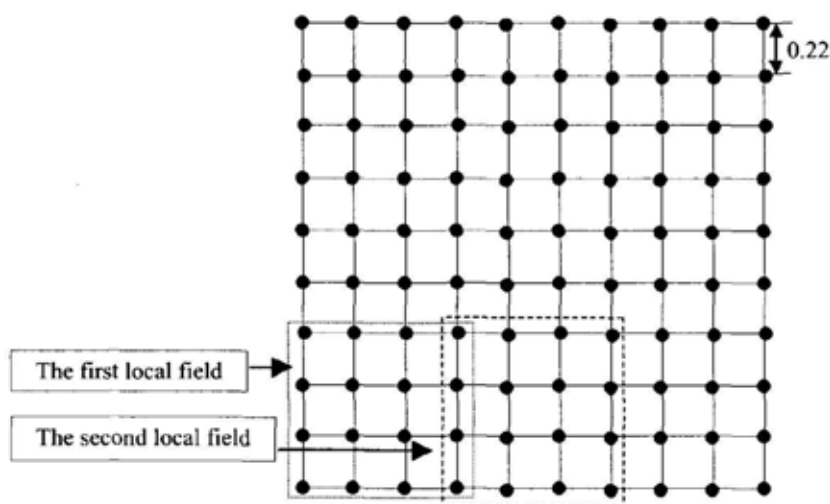


Figure 6. Sketch map of partition of the local field

#### 4.2 The way of selecting of vertical scaling factor

We give the way of selecting of vertical scaling factor as follows ( Hongquan Sun 1998):

① Using the practical data (interpolating data)  $\{x_i, y_j, z_{ij}\} (i=0, 1, \dots, N; j=0, 1, \dots, M)$  and based on the principle of the least square, we construct an one order trend surface equation:

$$\hat{z} = b_0 + b_1x + b_2y \quad (31)$$

② According to the one order trend surface equation, we calculate the trend value on each interpolating datum point:

$$\hat{z}_{i,j} = b_0 + b_1x_i + b_2y_j \quad (i = 0, 1, \dots, N; j = 0, 1, \dots, M) \quad (32)$$

③ Using the practical data to detract the trend value on each given point, we obtain the deviation value on the corresponding points:

$$e_{i,j} = z_{i,j} - \hat{z}_{i,j} \quad (i = 0, 1, \dots, N; j = 0, 1, \dots, M) \quad (33)$$

The relationships among given values, trend values and deviation values are shown in Figure 7.

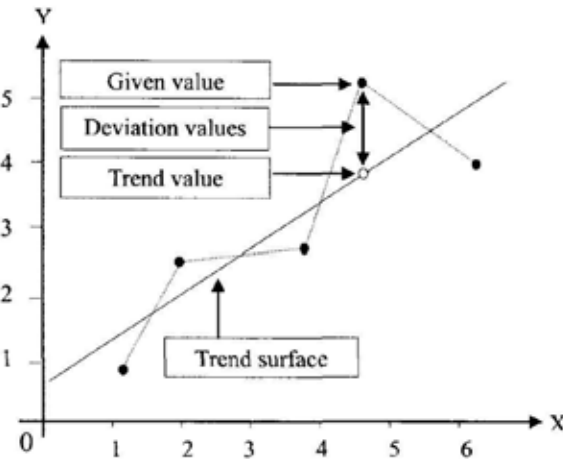


Figure 7. The relationship among given values, trend values and deviation values

#### 4.3 The deviation values can be used as the vertical scaling factors

In the fractal interpolation with the rectangle fields, the data on the regular gridding are expressed as  $\{x_i, y_j, z_{i,j}\} (i=0, 1, \dots, N; j=0, 1, \dots, M)$ . So the trend value and the deviation value are denoted by  $\hat{z}_{i,j}$  and  $e_{i,j}$  respectively. Then we have

Let:

$$e = \max_{0 \leq i \leq N; 0 \leq j \leq M} \{ |e_{i,j}| \} \quad (34)$$

Then the vertical scaling factors are found out by

$$s_{i,j} = e_{i,j} / e, \quad (i=0, 1, \dots, N; j=0, 1, \dots, M) \quad (35)$$

## 5. Fractal interpolation of the fault surface

Based on the method of fractal interpolation surface discussed above, the roughness of fractal surface is simulated by using the practical data shown in Figure 1. The result of the simulation is shown in Figure 8.

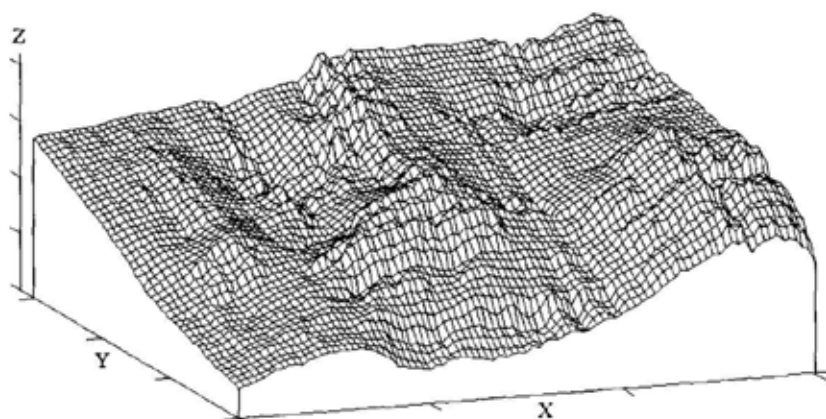


Figure 8. Fractal interpolated fault surface

From Figure 8, we can see that there are obviously scraggy local areas on the fractal interpolation fault surface. It gives us an intuitive roughness. It shows the virtue of the fractal interpolation. With the traditional methods, this result can't be obtained for the any closed points are connected by lines or smooth curves in traditional interpolation.

With the method of calculating box dimension (Hongquan Sun 1998), we can obtain that the dimension of fractal interpolated fault surface is 2.1993.

According to the principles of precision analyses (Hongquan Sun 1998), the dimension precision and the deviation precision of the fractal interpolated fault surface can be calculated. As the known data number is 28 and the simulated point number is 4186. So the information content is  $h = 28/4186 = 0.67\%$ . We obtain that the dimension precision and the deviation precision are 97.95% and 92.06% respectively.

## 6. Conclusions

Usually, the fracture surface in rocks appears to be statistically self-similar and/or self-affine. The fractal geometry supplies an alternative method to describe quantitatively the roughness of fault surfaces in geology. Extensive studies show that the morphology of fault surfaces in geology really affects the degree of the accidents of slope instability and roof caving in civil and mining engineering. However, for a geological fault or a joint, to obtain the morphology of whole fault surface is really difficult. It is to say that, in rock engineering, only a little information about roughness of a fault surface can be obtained from the exposure and several drill holes. The important work is how to estimate the morphology of fracture surfaces according to a little

information from the surface. Therefore, the theory and method of fractal interpolation is needed to develop a new way to estimate the morphology of an entire fault surface based on the information from drill holes.

In this paper, an improved self-affine fractal interpolation method is proposed. The variogram in geostatistics is introduced into the fractal interpolation of fault surfaces and the method of the local domain partition is established. By use of the principles of the trend surface analyses, the deviations on the information points are used as a vertical scaling factor.

The case studies presented in this paper demonstrate that the method of fractal interpolation surface is a very useful tool for simulation or generation of the morphology of fault surfaces.

### Acknowledgements

This work was supported by the Doctoral Researching Foundation of Hebei Province, P.R. China.

### References

1. Benoit B. Mandelbrot, *The fractal geometry of nature*: W. H. Freeman and Company, San Francisco: pp.361~366 (1982).
2. Heping Xie, et al, The influence of proximate fault morphology on ground subsidence due to extraction. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* 35(8),1107-1111 (1998).
3. Heping Xie and Hongquan Sun, The study on bivariate fractal interpolation functions and creation of fractal interpolated surfaces: *Fractals*, v. 5, no.4, pp. 625~634 (1997).
4. Hongquan Sun, *Geostatistics and its applications* (in Chinese): China University of Mining and Technology Press, pp. 59-89 (1990).
5. Hongquan Sun and Heping Xie, Fractal interpolation surface and its dimension estimation (in Chinese): *Journal of China University of Mining and Technology*, v. 27, no. 2, pp 217-220 (1998).
6. Hongquan Sun, The study on the theory of fractal interpolation surface and the interpolation of rock fracture surfaces (in Chinese): Doctoral Thesis, China University of Mining and Technology, Beijing, pp.60-92 (1998).
7. Kenneth Falconer, *Fractal geometry--Mathematical foundations and applications*: New York, pp.92-100 (1990).
8. Michael F. Barnsley, *Fractal functions and interpolations: Constructive Approximation* 2, p. 303-329 (1986).
9. Peter R. Massopust, *Fractal functions, fractal surfaces, and wavelets*: Academic Press, pp.135~355 (1994).

# A DETERMINISTIC POWER DOMAIN ALGORITHM FOR FRACTAL IMAGE DECOMPRESSION

N. NIKOLAOU AND A. KAKOS

*Division of Information Systems, Bank of Greece, 341 Mesogeion Ave., Cholargos 152  
32, Hellas*

*E-mail: npnikolaou@bankofgreece.gr; askakkos@bankofgreece.gr*

V. DRAKOPOULOS

*Department of Informatics and Telecommunications, Theoretical Informatics, University  
of Athens, Panepistimioupolis, Athens 157 84, Hellas*

*E-mail: vasilios@di.uoa.gr*

A new algorithm, called herein the Plotkin power domain algorithm, is discussed; it uses the Plotkin power domain as its computational model and it generates black and white images coded by an iterated function system, a technique used in fractal image compression. A simple complexity analysis for the algorithm is also derived.

## 1 Introduction

A number of algorithms have been proposed for the digitised approximation to the attractor of an (hyperbolic) *iterated function system*, or *IFS* for short, on the plane. Deterministic algorithms for decoding IFS-encoded-images involve determining all the descendants of seed pixels. In what follows, a competitive alternative will be described and implemented as a consequence of the introduction of Domain Theory in dynamical systems, measures and fractals<sup>1</sup>. It uses the Plotkin power domain as its computational model and generates black and white approximations to the attractors of various IFS's.

The proposed algorithm, after comparing with the most commonly used deterministic algorithms for the approximation of such attractors, namely the Deterministic Iteration Algorithm (DIA, see<sup>2</sup>), the Adaptive Cut Algorithm (ACA, see<sup>3</sup>) and the Minimal Plotting Algorithm (MPA, see<sup>4</sup>), shows to be, under all known circumstances, faster than DIA and ACA. Moreover, it can serve as a basis for a magnification algorithm, i.e. to render magnified fragments of fractal images; MPA shares with the DIA the defect that rendering a small part of a highly magnified attractor consumes inordinate amounts of memory.

An advantage of the new algorithm is that it encapsulates an economical stopping criterion; roughly speaking, we are in a position to know whether the actual attractor has been sufficiently produced in the space of the digitised screen. Moreover, for a given discrimination capability of the computer screen, the proposed algorithm has a determined upper and lower bound for the number of computations required before the best possible attractor for the given resolution is constructed. The exact number of computations, however, cannot be specified analytically in a closed formula, but can be very easily calculated with the aid of a computer.

An analytic description of the algorithm as well as a digest of the theoretical fundamentals, on which its model is based, follows. However, an extended abstract for the theory can be found in<sup>6</sup>, where power domains are discussed along with



IFS's.

## 2 A Computational Model

Although any complete metric space  $X$  would be sufficient as a model space for an (hyperbolic) IFS, in the case of computer graphics we are especially interested in the closed and bounded subsets of  $\mathbb{R}^2$ ; under these presuppositions our space  $X \subseteq \mathbb{R}^2$  is a compact metric space. Since the attractor of an IFS is a compact subset of  $X$ , it is natural to study this set as an element of the associated space  $\mathcal{U}X$ , namely the upper space of  $X$ , on which we will focus our attention.

For any Hausdorff metric space  $X$  the *upper space*  $\mathcal{U}X$  consists of all nonempty compact subsets of  $X$ , that is

$$\mathcal{U}X = \{\emptyset \neq C \subseteq X \mid C \text{ compact}\}.$$

This space has a topology, called the *upper topology*, whose base is the collection

$$\square_a = \{C \in \mathcal{U}X \mid C \subseteq a\},$$

where  $a \in \Omega X$  is an open set of  $X$ . This means that, for any open subset  $a$  of  $X$ , the collection of all compact nonempty subsets of  $X$  that are included in  $a$  forms an element of the base for the upper topology. This topology is  $T_0$ ; the specialisation ordering  $\sqsubseteq_u$  of  $\mathcal{U}X$  is the superset inclusion, i.e.

$$A \sqsubseteq_u B \iff \forall a \in \Omega X [A \subseteq a \Rightarrow B \subseteq a] \iff A \supseteq B.$$

Under this ordering ( $\sqsubseteq_u$ ) the space  $(\mathcal{U}X, \supseteq)$  becomes a *directed complete partial order* (d.c.p.o.), which means that every directed set has a *least upper bound* (l.u.b.). The l.u.b. of a directed set of compact subsets is their intersection; the elements of  $X$  are maximal elements of  $\mathcal{U}X$ .

Furthermore, whenever  $X$  is compact as in our case,  $(\mathcal{U}X, \supseteq)$  can be proved to be a bounded complete continuous d.c.p.o. and  $\mathcal{U}X$  has a bottom element. Since  $\mathbb{R}^2$  is second countable,  $X$  is a second countable space as well and the Proposition 3.4 in <sup>1</sup> suggests that  $X$  will have a countable basis of relatively compact neighbourhoods (i.e. their closures are compact sets) and that  $\mathcal{U}X$  will be  $\omega$ -continuous with an induced order basis consisting of finite unions of closures of these relatively compact neighbourhoods.

An (hyperbolic) IFS  $\{X; f_1, f_2, \dots, f_N\}$  or, more briefly,  $\{X; f_{1-N}\}$  induces a map  $F: \mathcal{U}X \rightarrow \mathcal{U}X$  defined by  $F(A) = f_1(A) \cup f_2(A) \cup \dots \cup f_N(A)$ , where  $f_i$  are contractions with corresponding contractivity factors  $s_i$  for  $i = 1, 2, \dots, N$ . Then  $F$  is a contraction with contractivity factor  $s = \max_i s_i$  and, according to the Contraction Mapping Theorem,  $F$  has a unique fixed point in  $\mathcal{U}X$ , which is called *the attractor* of the IFS.

We are able to construct the attractor as a result of a deterministic computation using the *Plotkin power domain*  $CUX$ , which contains the finite nonempty subsets of  $\mathcal{U}X$ , that is, all subsets of the form

$$a_I = \{a_i \in \mathcal{U}X \mid i \in I\}$$

for a finite nonempty indexing set  $I$ . Since  $\mathcal{U}X$  is  $\omega$ -continuous,  $CUX$  is an  $\omega$ -continuous d.c.p.o. and has an order basis consisting of equivalence classes of finite

sets  $\{A_1, A_2, \dots, A_k\}$  of basis elements  $A_i$  of  $\mathcal{UX}$  for  $1 \leq i \leq k$ , under the Egli-Milner preorder  $\ll_{EM}$ :

$$A \ll_{EM} C \text{ if and only if } (\forall a \in A \exists c \in C : a \ll c) \\ \text{and } (\forall c \in C \exists a \in A : a \ll c).$$

We denote this equivalence class by  $[\{A_1, A_2, \dots, A_k\}]$ . The IFS  $\{X; f_{1-N}\}$  induces a map  $\bar{F}: \mathcal{CUX} \rightarrow \mathcal{CUX}$  defined by  $\bar{F}([\{A_1, A_2, \dots, A_k\}]) = [\{f_i(A_j) \mid i = 1, 2, \dots, N, j = 1, 2, \dots, k\}]$ . The least fixed point of the function  $\bar{F}$  is  $\bigsqcup_{i \geq 0} \bar{F}^i([\{X\}])$  and can be obtained by constructing a finitely-branching tree as in Figure 1, which we call the *IFS tree* (descendant tree, tree of transformations or tree of images).

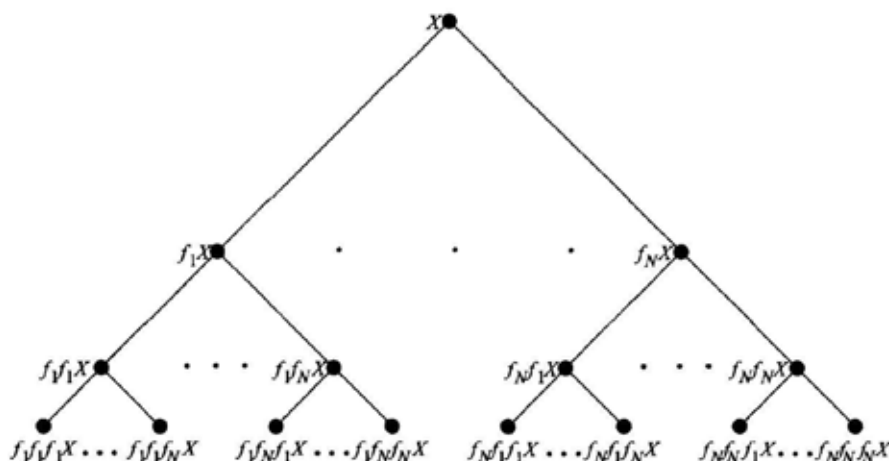


Figure 1. The IFS tree.

Since the affine transformations that we use are contractions, then for any branch at any depth  $n$  of the tree it will hold that  $\text{diam}(f_{i_1}f_{i_2} \dots f_{i_n}X) \leq s_{i_1}s_{i_2} \dots s_{i_n} \text{diam}(X) \leq s^n \text{diam}(X)$  and hence the l.u.b. of any infinite branch is a singleton set. The equivalence class of the set of l.u.b.'s of the infinite branches of the finitely-branching tree is the least fixed point of  $\bar{F}$  and hence these l.u.b.'s are exactly the points of the attractor of the hyperbolic IFS.

At this point, we have to note a very important property of the infinite finitely-branching tree, which is the basis for the new algorithm: every node of the tree is a compact set which includes all its children. For example, the node  $f_1X$  has the fundamental property that:

$$f_1X \supseteq f_1f_1X, f_1X \supseteq f_1f_2X, \dots, f_1X \supseteq f_1f_NX.$$

### 3 The Plotkin Power Domain Algorithm

We shall now study how the above mentioned theory can become an efficient algorithm for a conventional computing machine. In the following, the space  $X$  is

assumed to be the unit square, i.e.  $X = [0, 1] \times [0, 1]$ . This assumption will not harm the generality of the algorithm since any attractor can be 'moved' to the unit square under the use of appropriate transformations, such as shrinking, rotation or translation.

The transformations  $f_i, i = 1, 2, \dots, N$  used in an IFS are usually affine; affine transformations on  $\mathbb{R}^2$  have the very desirable property to map parallelograms to parallelograms. This means that, in our case, parallelograms will be produced at any branch at any level of the tree. This observation is a key element of the proposed algorithm as we shall see later on.

We saw that every branch of the infinite finitely-branching tree will yield a point of the attractor; nevertheless, the computer screen has only a limited resolution and, therefore, there is no need to go further down the tree than some level  $n$ . However, how can we determine which level is the one from which continuing onwards we obtain no more visible information for every branch of the tree? The answer to this question will be obtained after having computed the contractivity factor for every affine transformation of the IFS. For the following definition see also <sup>5</sup>.

**Definition 1** Let  $f$  be an affine transformation on  $\mathbb{R}^2$ , that is  $f\mathbf{x} = A\mathbf{x} + \mathbf{b}$ ,  $\mathbf{x}, \mathbf{b} \in \mathbb{R}^2$ ,  $A \in \mathbb{R}^{2 \times 2}$ ; define the norm  $\|A\|_2$  of the matrix  $A$  to be

$$\|A\|_2 = \max_{\|x\|=1} \|Ax\|, x \in \mathbb{R}^2.$$

The quantity  $\|A\|_2$  is the contractivity factor  $s$  for the affine transformation  $f$  since it holds that

$$\|f(x)\| \leq \|A\| \cdot \|x\|$$

and hence

$$\|f(x) - f(y)\| \leq \|A\| \cdot \|x - y\| \leq \|A\| \cdot \|x - y\|.$$

**Lemma 1** If the spectral radius of a matrix  $A \in \mathbb{R}^{n \times n}$  is defined to be

$$\rho(A) = \max \|\lambda\|, \lambda \in \mathbb{C}$$

where  $\lambda \in \mathbb{C}$  are the eigenvalues of  $A$ , then  $\rho(A^T A)^{1/2} = s = \|A\|_2$ .

The following lemma is an easy consequence of the above.

**Lemma 2** For a matrix

$$A = \begin{pmatrix} a & b \\ c & s \end{pmatrix} \in \mathbb{R}^{2 \times 2},$$

$$\|A\|_2 = \left( \frac{a^2 + b^2 + c^2 + s^2 + \sqrt{\Delta}}{2} \right)^{1/2},$$

where  $\Delta = ((a - s)^2 + (b + c)^2)((a + s)^2 + (b - c)^2) \geq 0$ ,  $\forall a, b, c, s \in \mathbb{R}$ .

*Proof.* Since

$$A = \begin{pmatrix} a & b \\ c & s \end{pmatrix},$$

then

$$A^T A = \begin{pmatrix} a^2 + c^2 & ab + cs \\ ab + cs & b^2 + s^2 \end{pmatrix}$$

is a symmetric matrix and the equation  $\det(A^T A - \lambda I) = 0$  will determine the two eigenvalues  $\lambda \in \mathbb{C}$  of  $A$ . However,  $\det(A^T A - \lambda I) = 0 \Leftrightarrow 0 = \lambda^2 - (a^2 + b^2 + c^2 + s^2)\lambda + [(a^2 + c^2)(b^2 + s^2) - (ab + cs)^2]$ . Since the quantity  $\Delta = ((a - s)^2 + (b + c)^2)((a + s)^2 + (b - c)^2) \geq 0$ , then  $A^T A$  has two distinct real eigenvalues  $\lambda \in \mathbb{R}$  whose  $\max\{|\lambda|, \lambda \in \mathbb{R}\}$  is

$$\max \left\{ \left| \frac{a^2 + b^2 + c^2 + s^2 + \sqrt{\Delta}}{2} \right|, \left| \frac{a^2 + b^2 + c^2 + s^2 - \sqrt{\Delta}}{2} \right| \right\} = \left\{ \frac{a^2 + b^2 + c^2 + s^2 + \sqrt{\Delta}}{2} \right\},$$

where  $\Delta \geq 0$ . □

By using the above formula we calculate exactly the contractivity factors  $s_1, s_2, \dots, s_N$  for each of the affine transformations  $f_1, f_2, \dots, f_N$  of the system. At the root of the tree lies the space  $X$ , which in our case is the unit square  $[0, 1] \times [0, 1]$  whose diameter is  $\sqrt{2}$ ; at the first level every parallelogram  $f_{i_1} X$  has a corresponding diameter less than or equal to  $s_{i_1} \sqrt{2}$ , at the second level the parallelogram  $f_{i_1} f_{i_2} X$  has a corresponding diameter less than or equal to  $s_{i_1} s_{i_2} \sqrt{2}$ . At the  $n$ -th level the parallelogram  $f_{i_1} f_{i_2} \dots f_{i_n} X$  has a corresponding diameter less than or equal to  $s_{i_1} s_{i_2} \dots s_{i_n} \sqrt{2}$ . It is obvious that, if at some level  $n$  the diameter of a parallelogram becomes less than  $\varepsilon = 1/M$ ,  $M$  being the resolution of the screen, then there is no need to go further downwards to other levels for, the image produced will not have any noticeable improvement.

In other words, we construct all branches of the tree, until that level  $n$  for which the following condition becomes true:  $s_{i_1} s_{i_2} \dots s_{i_n} \sqrt{2} < \varepsilon$ . Since the transformations  $f_i$  are contractions ( $s_i < 1$ ), then the quantity  $s_{i_1} s_{i_2} \dots s_{i_n} \sqrt{2}$  will be strictly decreasing as  $n$  increases and hence for any branch of the tree there will be some  $n \in \mathbb{N}$  so that this condition will be satisfied. This proves that the algorithm is bound to terminate.

We just saw that every node of the tree is a parallelogram (subset of  $[0, 1] \times [0, 1]$ ) and that each branch converge to a point of the attractor which, for the case of the computer screen, can be computed in a finite number of steps. Hence we have a sequence of parallelograms  $\{A_n \mid n \in \mathbb{N}\}$  which converge to a point; since we are interested in the computation of this point and not in the computation of the intermediate parallelograms, it is rational not to compute - at each step - the parallelogram  $A_i$  produced, but any point  $a_i$  within this parallelogram  $A_i$ . Since the decreasing sequence of parallelograms  $\{A_n \mid n \in \mathbb{N}\}$ ,  $A_n \subset A_{n-1}$  converges to a point  $a$ , then every sequence of points  $\{a_n \mid a_n \in A_n, n \in \mathbb{N}\}$  which are contained in these parallelograms will eventually converge to the same point  $a$ . This means that if  $\lim_{n \rightarrow \infty} A_n = \{a\}$  and  $a_n \in A_n, \forall n \in \mathbb{N}$ , then the sequence  $\{a_n \mid n \in \mathbb{N}\}$  will converge to  $a$  for any choice  $a_n$  out of the corresponding parallelogram  $A_n$ . This observation leads to a significant improvement in the runtime required to construct the attractor; at each step only two points have to be calculated.

Furthermore, an efficient implementation should not make any computation more than once; this means that  $f_1X, f_2X, \dots, f_NX, f_1f_1X, f_1f_2X, \dots, f_Nf_NX, \dots$  should be computed only once. It is more than obvious, however, that all the nodes of this immense tree cannot be stored in the computer memory. Since it is not feasible to avoid redundant computations when we use the above tree, in our implementation the tree of Figure 2 is used which, for computational purposes, it is proved to be equivalent.

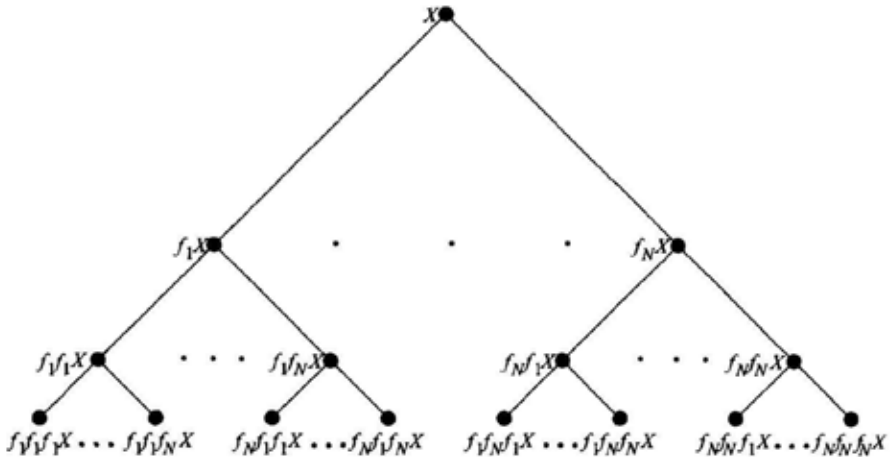


Figure 2. The action tree.

The space  $X$  stands at the root of the tree; its children,  $f_1X, f_2X, \dots, f_NX$ , will be the image of  $X$  under the affine transformations of the IFS. In general, the node  $f_{i_1}f_{i_2} \dots f_{i_{N-1}}f_{i_N}X$  will have  $f_{i_N}f_{i_1}f_{i_2} \dots f_{i_{N-1}}f_1X, f_{i_N}f_{i_1}f_{i_2} \dots f_{i_{N-1}}f_2X, \dots, f_{i_N}f_{i_1}f_{i_2} \dots f_{i_{N-1}}f_NX$  as its children.

This equivalent tree is our main contribution to the implementation of the algorithm. Although the algorithm incorporates a lot of new ideas, without that tree its efficient implementation would have been out of question. Alternatively, one should either have to store all the nodes of the tree in the computer memory or some computations would have to be performed more than once. In the first case, however, the implementation would not be optimal in terms of memory usage, whereas in the second case the speed of the algorithm would decrease substantially.

The two trees have exactly the same first two levels. From the third level onwards, however, the differences start. Even if the second level is identical in both trees, a different third level is produced. It is our intention to prove that this does not affect the theoretical basis and that the third level is computationally equivalent. If at some node of the second level holds that  $s_{i_1}s_{i_2}\sqrt{2} < \varepsilon$  then, due to the commutative property of the real numbers, also holds that  $s_{i_2}s_{i_1}\sqrt{2} < \varepsilon$ . Hence, in the case of the first tree,  $s_{i_1}s_{i_2}\sqrt{2} < \varepsilon$  means that the children of  $f_{i_1}f_{i_2}X$  (that is  $f_{i_1}f_{i_2}f_1X, f_{i_1}f_{i_2}f_2X, \dots, f_{i_1}f_{i_2}f_NX$ ) will not be produced and the pixel  $f_{i_1}f_{i_2}X$  will be plotted. For this same tree,  $s_{i_2}s_{i_1}\sqrt{2} < \varepsilon$  means that

the children of  $f_{i_2} f_{i_1} X$  (that is  $f_{i_2} f_{i_1} f_1 X, f_{i_2} f_{i_1} f_2 X, \dots, f_{i_2} f_{i_1} f_N X$ ) will not be produced and the pixel  $f_{i_2} f_{i_1} X$  will be plotted. On the other hand, for the second tree, the condition  $s_{i_1} s_{i_2} \sqrt{2} < \varepsilon$  means that the children of  $f_{i_1} f_{i_2} X$  (that is  $f_{i_1} f_{i_2} f_1 X, f_{i_1} f_{i_2} f_2 X, \dots, f_{i_1} f_{i_2} f_N X$ ) will not be produced and the pixel  $f_{i_1} f_{i_2} X$  will be plotted. Moreover, the condition  $s_{i_2} s_{i_1} \sqrt{2} < \varepsilon$ , for the second tree, means that the children of  $f_{i_2} f_{i_1} X$  (that is  $f_{i_1} f_{i_2} f_1 X, f_{i_1} f_{i_2} f_2 X, \dots, f_{i_1} f_{i_2} f_N X$ ) will not be produced and the pixel  $f_{i_2} f_{i_1} X$  will be plotted. It is now more than obvious that the third level of the two trees are computationally equivalent. Using induction we can prove that the two trees are equivalent at all levels. Thus the two trees are, for our purposes, computationally equivalent.

The second tree, however, has the advantage that in order to compute the nodes of a level it suffices to store only  $N$  values of the nodes of the previous level. The use of the second tree enables us to avoid all unnecessary recomputations; actually not even a single recomputation is performed. Furthermore, a tremendous storage economy is achieved since at level  $n$  we store only  $N$  out of  $N^n$  values of that level. This has the additive effect that for the generation of the  $(n+1)$  level only  $n \cdot N$  values, instead of  $N + N^2 + \dots + N^n = N(N^n - 1)/(N - 1)$ , need be stored.

#### 4 The Algorithm and its Complexity

The actual algorithm in a form of pseudocode, which also provides a definition for the equivalent tree of Figure 2, has as follows:

0. Start.
1. Compute all contractivity factors  $s_i$ .
2. Call **procedure produce**  $([x_0, x_0, \dots, x_0], 0, [\sqrt{2}, \sqrt{2}, \dots, \sqrt{2}])$ , for  $x_0 \in [0, 1]^2$ .
3. End.

where

```

procedure produce  $([x_1, x_2, \dots, x_N], q, [d_1, d_2, \dots, d_N])$  {
  for  $i = 1, \dots, N$  do {
     $x'_i = f_i(x_q)$ 
     $d'_i = s_i * d_q$ 
  }
  for  $i = 1, \dots, N$  do {
    if  $(d'_i > 1/M)$  then do
      call produce  $([x'_1, x'_2, \dots, x'_N], i, [d'_1, d'_2, \dots, d'_N])$ 
    else do
      plot pixel  $x'_i$ 
    }
  }
}

```

Having described the algorithm, we shall try to identify the number of computations needed for the construction of the attractor. Since the contractivity factors of the affine transformations are known, we can find an estimation for the depth

Table 1. The number of pixels drawn for the systematic comparison.

Pixels drawn	Plotkin PDA	ACA	DIA	MPA
Sierpiński	6561	6561	2403	2020
Dendrite	8617	12349	1329	1231
Spiral	3464	4838	940	481
Meander	6561	6561	1692	1507
Dendrite II	5413	6561	1992	1860
Fern leaf	57031	71776	1355	1168
Maple leaf	206080	206080	2191	2184
Eye	380881	535533	1174	1164

of the tree. If  $s_{max}$  and  $s_{min}$  are the largest and the smallest contractivity factors, respectively, then the depth of the tree will lie between

$$d_{max} = \left\lceil \frac{-\ln M}{\ln s_{max}} \right\rceil + 1 \text{ and } d_{min} = \left\lceil \frac{-\ln M}{\ln s_{min}} \right\rceil + 1$$

that are the greatest and the smallest depth, respectively. This means that in the worst case we need  $9(N + N^2 + \dots + N^{d_{max}}) = 9[(N^{d_{max}+1} - 1)/(N - 1) - 1]$ , thus giving  $O(N^{d_{max}})$  while in the best case we need  $9(N + N^2 + \dots + N^{d_{min}}) = 9[(N^{d_{min}+1} - 1)/(N - 1) - 1]$  computations. This number is explained by the fact that 8 computations are needed for the calculation of the new point and only one computation is needed for the calculation of the quantity  $s_{i_1}(s_{i_2} \dots s_{i_n} \sqrt{2})$  since at each step the quantity  $s_{i_2} \dots s_{i_n} \sqrt{2}$  has already been computed in previous nodes.

The exact number of computations performed before the construction of the attractor is  $C(\sqrt{2})$ , where

$$C(x) = \begin{cases} 0, & \text{if } x < \varepsilon \\ \sum_{i=1}^N C(s_i x) + 9N, & \text{otherwise,} \end{cases}$$

where  $N$  is the number of affine transformations and  $M$  is the resolution of the screen. Although, it is extremely difficult to find a closed formula which gives the exact number of computations, the above mentioned recursive formula can be used to obtain this number with the aid of a computer.

## 5 Conclusions

The current implementation of our algorithm is written in Microsoft Visual Basic 6.0. It is capable of drawing fractal images using the Plotkin PDA, the DIA, the ACA and the MPA, and displaying the depth of the action tree, the number of points used for rendering and the total runtime. The fractal images ( $M = 100$ ) used for the comparisons of the various algorithms are illustrated in Figure 3. The Plotkin PDA was finally tested and rated by comparing the various fractal attractors produced by it versus the attractors produced using the DIA (level=30), the ACA and the MPA. Time results are given in CPU seconds on a Pentium IV PC with a 1.5 GHz CPU clock running Windows 2000 SP 3.

As can be seen from Figure 4, our algorithm is extremely efficient for the decoding of pictures which have been compressed using some method of fractal



Figure 3. The Sierpiński triangle, a dendrite, a spiral, a second dendrite, a meander, a fern leaf, a maple leaf and an eye.

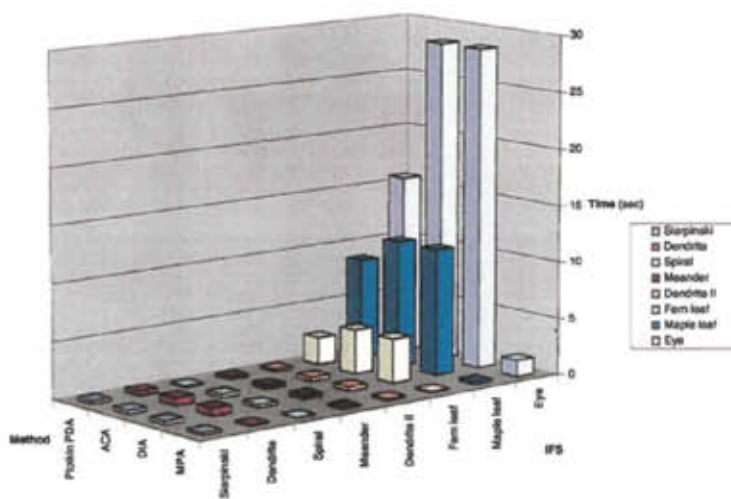


Figure 4. The time results of the systematic comparison.



compression based on IFS. This is explained by the fact that natural images - such as the face of a person - lack self-similarity and hence a lot of affine transformations are employed for the coding. Each of them, however, will have a very small contractivity factor and hence our tree will have small depth (10 for the maple leaf, 31 for the fern leaf, 45 for the spiral, 3 for the eye and 8 for the other figures); therefore, the construction of the attractor will terminate very quickly, except for the maple leaf, since the Plotkin PDA as well as the ACA ignore overlapping. The number of points drawn are shown in Table 1.

## Appendix

We list in Tables 2-8 the IFS codes for some of the examples discussed in the main text.

Table 2. The IFS code for the Sierpiński triangle.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0.5	0	0	0.5	0	0
2	0.5	0	0	0.5	0.5	0
3	0.5	0	0	0.5	0.25	0.433

Table 3. The IFS code for a dendrite.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0.5	0	0	0.5	0.0625	0.15
2	0.21	-0.20625	0.528	0.21	0.789	0
3	0.5	0	0	0.5	0.375	0.375
4	-0.2	0.1125	-0.288	-0.2	0.609	0.975

Table 4. The IFS code for a spiral.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	-0.18	0.126	-0.2571	-0.18	0.815	0.8485
2	-0.8	0.4	-0.4	0.8	-0.088	0.2514

Table 5. The IFS code for a second dendrite.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0.5	0	0	0.5	0.3125	0
2	0.5	0	0	0.5	0.496094	0
3	0.28	-0.25	0.64	0.28	0.523437	0.05

Table 6. The IFS code for a meander.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0	0.33	-0.768	0	0.4609	0.9375
2	0.5	0	0	0.22	0.5	0.9375
3	0.5	0	0	0.24	0.5	0.9375

Table 7. The IFS code for a fern leaf.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0	0	0	0.16	0.5	0.07
2	0.2	-0.195	0.3066667	0.22	0.41625	0.045
3	-0.15	0.21	0.3466667	0.24	0.5575	-0.07333
4	0.85	0.03	-0.5333	0.85	0.07249999	0.1725

Table 8. The IFS code for a maple leaf.

$f$	$a$	$b$	$c$	$s$	$d$	$e$
1	0.6	0	0	0.6	0.18	0.36
2	0.6	0	0	0.6	0.18	0.12
3	0.4	0.3	-0.3	0.4	0.27	0.36
4	0.4	-0.3	0.3	0.4	0.27	0.09

## References

1. A. Edalat, *Inform. and Comput.* **120**, 32 (1995).
2. M.F. Barnsley, *Fractals everywhere*, 2nd ed., (Academic Press Professional, San Diego, 1993).
3. D. Hepting, P. Prusinkiewicz and D. Saupe in *Fractals in the fundamental and applied sciences*, eds. H.-O. Peitgen, J.M. Henriques and L.F. Penedo (North-Holland, Amsterdam, 1991).
4. D.M. Monro and F. Dudbridge, *IEEE Comput. Graphics Appl.* **15**, 32 (1995).
5. R.L. Burden and J.D. Faires, *Numerical Analysis*, 4th ed., (PWS Publishing Company, Boston, 1989).
6. A. Edalat, *Inform. and Comput.* **124**, 182 (1996).

This page intentionally left blank

# COMPARATIVE DYNAMIC SCALING ANALYSIS OF QUASI-2D ELECTRODEPOSITED SILVER PATTERNS UNDER LOCALIZED AND NON-LOCALIZED RANDOM QUENCHED NOISE

M. A. PASQUALE

*Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA, (Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad Nacional de La Plata),  
Sucursal 4, Casilla de Correo 16,(1900) La Plata, Argentina  
e-mail: miguel@inifta.unlp.edu.ar*

S. L. MARCHIANO

*Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA, (Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad Nacional de La Plata),  
Sucursal 4, Casilla de Correo 16,(1900) La Plata, Argentina  
e-mail: smarchiano@inifta.unlp.edu.ar*

A. J. ARVIA

*Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA, (Consejo Nacional de Investigaciones Científicas y Técnicas-Universidad Nacional de La Plata),  
Sucursal 4, Casilla de Correo 16,(1900) La Plata, Argentina  
e-mail: ajarvia@inifta.unlp.edu.ar*

Data from the dynamic scaling analysis of the growth front of silver patterns electroformed in a quasi-2D cell under localized and non-localized random quenched noise are reported. The plating solution either embedded in filter paper (FP), or containing disordered glass beads (GB), or as agarose gels (AG) were utilized. The scaling exponents from the displacement of the driven interface are  $\alpha = 0.63 \pm 0.05$  and  $\beta = 0.60 \pm 0.05$  for FP, irrespective of its pore size distribution;  $\alpha = 0.64 \pm 0.05$  and  $\beta = 0.58 \pm 0.05$  for GB; and  $\alpha = 1.25 \pm 0.10$  and  $\beta = 0.88 \pm 0.15$  for AG. Exponents for FP and GB fit the predictions of the directed percolation depinning (DPD) model for  $D = 1$ , whereas for AG they coincide with those calculated by Leschhorn from a lattice model of probabilistic cellular automata. The difference between exponents resulting from FP, GB, and AG can be attributed to a non-localized random pinning in AG, which introduces a size-dependence mobility of obstacles in the gelled medium.

Keywords: Pinning, Directed Percolation Depinning, Cellular Automata.

## 1 Introduction

The behavior of driven interfaces produced far from equilibrium and subjected to quenched random forces remains a challenging problem. These driven interfaces may exhibit self-similar, self-affine or even non-fractal behavior. The driven force ( $F$ ) may result from pressure gradients, magnetic fields as in the ordering kinetics of impure magnets<sup>1,2</sup>, and a chemical potential favoring the growth of one of the coexisting phases. For  $F < F_c$ , where  $F_c$  is a critical value of  $F$ , the interface is pinned and its front displacement is only possible due to thermal fluctuations, whereas, for  $F > F_c$ , the interface is capable of moving with a finite velocity. Accordingly,  $F_c$  is related to a pinning-depinning transition. Considerable progress

has been made in understanding the dynamics of non-equilibrium interface growth in a medium with random pinning forces in the context of experiments, analytical theories, and a variety of models<sup>3</sup>. These experiments included, among others<sup>4</sup>, the growth of bacterial colonies<sup>5</sup>, fluid flow displacement experiments in porous media<sup>6,7,8,9</sup>, paper<sup>10</sup> and sponge-like material<sup>11</sup> wetting, propagation of burning fronts<sup>12</sup>, and paper tearing<sup>13</sup>. Plausible continuum descriptions of the interface dynamics under a uniform random driving force  $F$ , based essentially on a roughening and smoothing mechanism, are given either by the Langevin equation<sup>14,15</sup>:

$$\frac{\partial h}{\partial t} = \nu \nabla^2 h + F - \eta(x, h). \quad (1)$$

or the Kardar-Parisi-Zhang (KPZ)<sup>16</sup> non linear equation:

$$\frac{\partial h}{\partial t} = \nu \nabla^2 h + \frac{\lambda}{2} (\nabla h)^2 + F - \eta(x, h). \quad (2)$$

where  $h = h(x, t)$  is the height of the interface at position  $x$  at time  $t$ .

The first term in equations (1) and (2) accounts for the growth front smoothing by surface tension-like relaxation,  $\nu$  being a surface tension coefficient. The second term in equation (2) accounts for the directional velocity of impinging particles referred to the tangent plane of the growing front, being  $\lambda$  a constant. Other terms in equations (1) and (2) contribute to roughening. The term  $\eta(x, h)$  represents a quenched random force rather than fluctuating in time as in the original Edwards-Willkinson or the Kardar, Parisi and Zhang (KPZ) equations. Numerical solutions of equations (1) and (2) have been obtained under the assumptions that the noise term  $\eta(x, h)$  can be expressed as a power law correlation<sup>17</sup> the existence of long-range correlated noise<sup>18,19,20</sup>, and hinting the role of anisotropy<sup>21</sup> in order to reproduce experimental data. Several attempts have also been made to understand the roughening behavior of the interface on a phenomenological level by analyzing models that incorporate the essential roughening and smoothing mechanism<sup>22</sup>. Models based on the directed percolation theory<sup>23</sup> have been applied successfully in many situations<sup>10,24</sup> and extended to others with the incorporation of several modifications<sup>25,26</sup>. Nevertheless, despite these advances, the physical meaning of results from either lattice models or continuum equations, is not fully agreed because of the large dispersion of experimental data<sup>3</sup>. We report the experimental scaling exponents resulting from the dynamics of quasi-2D silver electrodeposits made in different aqueous environments under either localized or non-localized random field pinning. Exponents for filter paper and disordered glass beads fit the predictions of the directed percolation depinning (DPD) model for  $D = 1$ . Experimental exponents for agarose gels show an anomalous roughness in comparison to the predictions of the KPZ equation that can be explained by the lattice model of probabilistic cellular automata.

## 2 Experimental

Silver patterns were grown utilizing a linear quasi 2D electrochemical cell with a parallel plate silver anode and cathode arrangement at 2 cm distance. Further

details of the electrochemical cell design were described elsewhere<sup>27</sup>. Two different plating solutions (I and II) were utilized. Solution I consisted of  $x$  M silver perchlorate + 1 M perchloric acid ( $0.005 \text{ M} < x < 1 \text{ M}$ ). This solution was used either embedded in filter paper (FP) of different pore size, i.e., Whatman 41\* and Whatman 42\* or in a bed of glass beads  $5 \mu\text{m}$  average diameter (GB). Solution II was an agarose gel (AG) made of  $y$  M silver sulfate + 0.5 M sodium sulfate + 0.01 M sulfuric acid + 0.6% w/v agarose ( $0.008 \text{ M} < y < 0.024 \text{ M}$ ). The AG was prepared by first dissolving silver sulfate and sodium sulfate in hot water, then adding agarose and heating up to the boiling point, and finally, while cooling, sulfuric acid was added. The hot agarose sol was then poured into the cell and cooled down for gelling. For bath plating preparation analytical quality reagents and Milli-Q\* water were employed. Disordered media were placed between the cathode and anode, and pressed between Lucite\* or glass plates at a 0.025 cm distance. Silver patterns were grown under a constant cathode-to-anode potential in the range  $-1.0 < \Delta E_{ac} < -1.2 \text{ V}$  utilizing a Radiometer 320 potentiostat. For this range of  $\Delta E_{ac}$  the electrochemical reaction was under a mass transport regime as was concluded from the cathodic polarization curve. For each run, the cathodic current ( $I_c$ ) and charge ( $Q_c$ ) transients were recorded. The interface motion was followed utilizing a charge-coupled device video camera (Hitachi 220) coupled to a stereoscopic microscope (C. Zeiss Stemi 200). The images were digitized with a spatial resolution of  $568 \times 744$  pixels and 8-bit intensity resolution using a KS 300 Kontron Electronics frame grabber. The imaging system was arranged to capture only the central 1 cm wide domain of the interface to avoid edge effects.

### 3 Results and Discussion

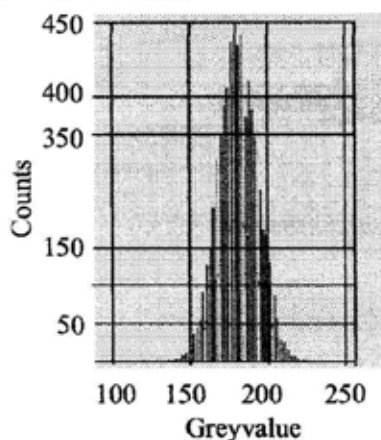
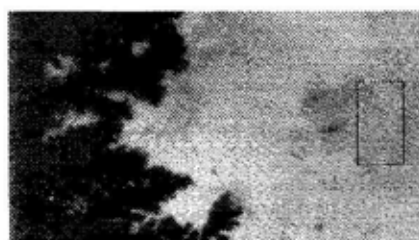
We found a remarkable difference in the growth pattern dynamics resulting from runs made in FP, GB, and AG media, as the localized random quenched noise for FP and GB contrasts with the random quenched noise resulting from AG. In fact, the AG structure consists of random distributed agglomerates of different size, covering the range from a fraction of a  $\mu\text{m}$  to approximately  $1000 \mu\text{m}$ . The average channel width between agglomerates is near  $400 \mu\text{m}$ , a figure that is close to the average branch width of silver electrodeposits produced in AG. It is worth to realize that the gel structure is very sensitive to the gel preparation procedure<sup>28</sup>. For such a size distribution of agglomerates the possibility exists that small agglomerates move faster than larger ones. The percolation of the largest agglomerates becomes also possible, approaching localized random quenching. Therefore, in contrast to FP and GB, the size and velocity distributions of agglomerates in AG would represent a new random pinning situation that we refer to as either non-localized or pseudo-quenched random noise. Thus, for AG, the description of pinning forces would result more complicated than for FP and GB. The validity of the agglomerate motion hypothesis in AG was demonstrated by running the following experiment. First, a fringe of approximately 0.5 cm of AG gel was placed in contact with the cathode surface. Then, a portion of methylene blue-stained colored AG was added between the fringe and the counter electrode, and finally pressed between the plates of the cell. This resulted in a predominantly uncolored

AG fringe on the cathode surface and AG with diffuse colored spots filling the rest of the cell. In a black and white digitized image, a 200 x 700 mm reference box was drawn at the border of a spot of methylene blue-stained gel opposite to the cathode. A first gray scale histogram (blank) was constructed for the pixels inside the box. Then, after the growing front was just going to touch the reference box edge, another gray scale histogram was made. This histogram showed the appearance of a larger number of darker pixels in the reference box (Fig. 1), as expected from the displacement of small agarose agglomerates pushed ahead by the moving front. Therefore, agglomerates, particularly those of smaller size, should contribute as moving obstacles for making the displacement of the growing front more difficult. The sieving effect of agarose in the conducting medium would be equivalent to pinning the rate of motion of the growth front at two length and time scales that are associated with the random distribution of agglomerates and channels, and the appearance of a pseudo pinning/depinning transition<sup>28</sup>. Silver growth patterns formed at  $\Delta E_{ac} = -1.20$  V from agarose-free 0.024 M aqueous silver sulfate + 0.5 M sodium sulfate + 0.01 M sulfuric acid consist of a first rather compact layer about 0.06 mm thick produced for  $0 < t < 15$  s, followed by a branched layer about 0.21 mm thick, and later by a dense branched layer (Fig. 2a). On the other hand, growth patterns run in 0.6% w/v agarose gel exhibit the first rather compact layer for  $0 < t < 30$  s followed by a densely packed branching for  $30 < t < 200$  s, and finally, a small number of columns with a fan-like dense branching (Fig. 2b). We note that these changes cannot be attributed to any possible interference of agarose in the proper electron transfer process. Its presence only produces a decrease in the cathodic limiting current by about 25% in going from agarose-free solution to gels<sup>28,29</sup>, which is attributed to the sieving effect of the medium. At any rate, the mobility of small agglomerates has to be considered in relation to the impinging rate of discharging ions on the surface of the growing phase. For a mass transport limiting current regime, the impinging rate depends on the local concentration gradient of silver cations at the growing front and, therefore, can be modified by adjusting the concentration of these ions in the medium. For the analysis of the above pseudo-noise, one can define  $\langle V_f \rangle$ , the average directional growing front velocity,  $V_p$ , the velocity of silver cations in the medium that is related to the diffusion coefficient of these ions, and  $V_L$ , the random walk velocity of agarose agglomerates. For large agglomerates  $V_L \rightarrow 0$ , and for smaller ones  $V_L$  increases with the reciprocal of the average agglomerate size. Thus, for  $\langle V_f \rangle \gg V_p < V_L$  the growing front displacement is under pinning, whereas depinning occurs for  $\langle V_f \rangle \gg V_p \cong V_L$  (Fig. 3).

The morphology of the driven interface obtained in FP and GB shows comparable sieving effects (Fig. 2c-e) such as those already described for AG, despite the fact that FP and GB involve localized random pinning obstacles. For the same electrodeposition time ( $t$ ), the comparison of growth patterns run with Whatman 42\* (Fig. 2c) and Whatman 41\* (Fig. 2d) shows a slower front displacement for the enhanced pinning forces in Whatman 42\*. Similar results have been reported for glass bead fluid flow experiments<sup>7</sup>, in which both the permeability  $k$  and the fluid displacement velocity decrease by using smaller size glass beads.

The dynamic scaling theory predicts that  $W(L, t)$ , the interface width of a grow-

(a)



(b)

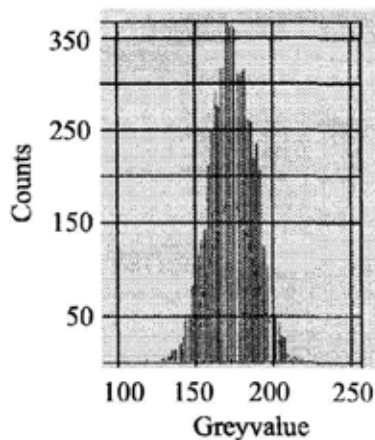
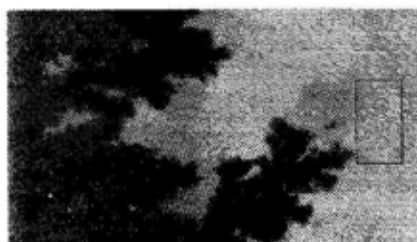


Figure 1. Images of silver quasi-2D growth patterns formed from 0.024 M silver sulfate + 0.5 sodium sulfate + 0.01 M sulfuric acid + 0.6% w/v agarose gel at  $\Delta E_{c-a} = -0.120$  V and 298 K, and their corresponding histograms. A reference box was drawn at the border of a spot of methylene blue-stained gel opposite to the cathode. The gray scale histogram shown in (a) was obtained either in the absence of silver electrodeposit or when the latter was still far from the reference box edge. The histogram shown in (b) was obtained when the electrodeposit front was just going to touch the reference box edge. In this case, the gray density (darker pixels) at the left-hand side of the reference box has notoriously increased, and then the histogram shows a significant change in half width, symmetry, as well as a shift of its maximum value. These effects are attributed to the motion of small size stained gel agglomerates.



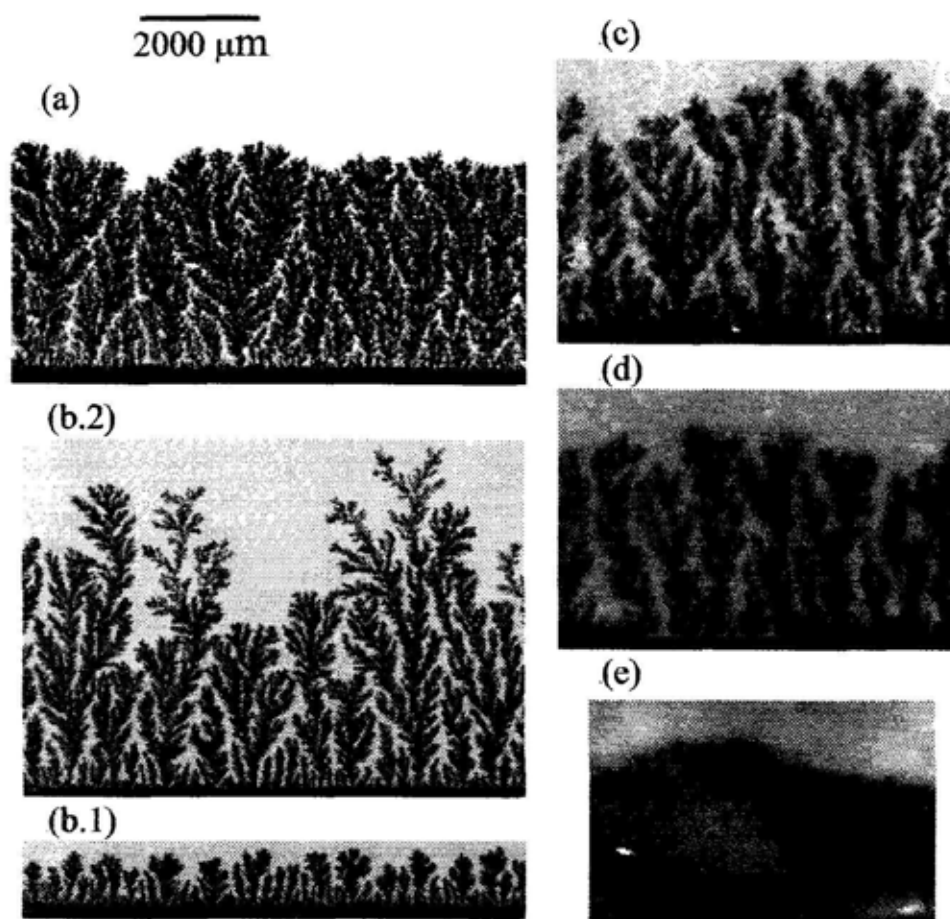


Figure 2. Quasi-2D silver patterns obtained at  $\Delta E_{c-a} = -0.120$  V and 298 K, from different media from (a) 0.024 M silver sulfate + 0.5 sodium sulfate + 0.01 M sulfuric acid at  $t = 300$  s; (b) 0.024 M silver sulfate + 0.5 sodium sulfate + 0.01 M sulfuric acid + 0.6% w/v agarose (gel) at  $t = 120$  s (b.1),  $t = 1020$  s (b.2); (c) 1 M silver perchlorate + 1 M perchloric acid embedded in filter paper Whatman 41\*,  $t = 480$  s; (d) 1 M silver perchlorate + 1 M perchloric acid embedded in filter paper Whatman 42\*,  $t = 480$  s; (e) 1 M silver perchlorate + 1 M perchloric acid in 5 μm glass beads,  $t = 300$  s.

ing front at time  $t$  and length scale  $L$ , is

$$W(L, t) = \sqrt{\frac{1}{L} \sum_{i=1}^L [h(i, t) - \bar{h}(t)]^2} \quad (3)$$

where  $h(i, t)$  is the height of the  $i$  column, and the mean height is

$$\bar{h}(t) = \frac{1}{L} \sum_{i=1}^L h(i, t) \quad (4)$$

for length scale  $L$  and growing time  $t$ , equation(3) scales as <sup>30</sup>

$$W(L, t) \propto L^\alpha f(t/L^{\frac{\alpha}{\beta}}) \quad (5)$$

Equation (5) for  $t \gg L^{\frac{\alpha}{\beta}}$  becomes  $W(L, t) \propto L^\alpha$ , and for  $t \ll L^{\frac{\alpha}{\beta}}$  it becomes  $W(L, t) \propto t^\beta$ . Then, from these equations the roughness exponent  $\alpha$  and the growth exponent  $\beta$  can be evaluated. The value of  $\alpha$  is related to the surface texture and to the fractal surface dimension  $D_F$  of the self-affine surface,  $D_F = 3 - \alpha$ . Thus, for  $\alpha \rightarrow 1$ ,  $D_F \rightarrow 2$ , i.e, the surface tends to be Euclidean (ordered), while, for  $\alpha \rightarrow 0$ ,  $D_F \rightarrow 3$ , the surface exhibits an increasing degree of disorder (fractal). The exponents  $\alpha$  and  $\beta$  are not independent and there is a simple way to collapse the temporal and spatial  $W(L, t)$  data onto a single curve by plotting  $W/L^\alpha$  versus  $t/L^{\frac{\alpha}{\beta}}$ .

Experimental data plotted as  $\log W(L, t) = \log W_{rms}$  versus  $\log t$  and  $\log W(L, t) = \log W_{rms}$  versus  $\log L$  after saturation time ( $t \gg L^{\frac{\alpha}{\beta}}$ ) from runs made in AG (Fig. 4.3), FP (Fig 4.1), and GB (Fig. 4.2) exhibit reasonable linear regions approximately one order of magnitude in both axes. These plots show, however, different functionalities that depend on the characteristics of the medium <sup>24,25,26,29</sup>.

Scaling exponents resulting from runs made in pinning-free medium (Fig. 2a) are  $\alpha = 0.5$  and  $\beta = 0.33$ , in agreement with KPZ equation as earlier reported <sup>27,31</sup>. Data from our disordered media are  $\alpha = 0.63 \pm 0.05$  and  $\beta = 0.60 \pm 0.05$  for FP, irrespective of its pore size distribution;  $\alpha = 0.64 \pm 0.05$  and  $\beta = 0.58 \pm 0.05$  for GB; and  $\alpha = 1.25 \pm 0.10$  and  $\beta = 0.88 \pm 0.15$  for AG. Despite data scattering, there is a significant difference between the three media. The values of the scaling exponents obtained from runs made with both FP and GB agree with the predictions of the directed percolation depinning (DPD) model <sup>9,24,25</sup>. On the other hand, results from gels are consistent with a lattice model based on cellular automata <sup>32</sup>.

The value of  $\alpha$  in AG is much larger than that expected from equation (1). This would mean that quenched noise roughens the interface more than thermal noise. On the other hand, the value of  $\beta$  above 0.5 points out the development of an unstable interface. Finally, the value of the dynamic exponent,  $z = \frac{\alpha}{\beta} = 1.42$  indicates that the dynamics of the depinning transition is superdiffusive. Thus, the scaling exponents show a remarkable dependence on whether the random pinning forces are localized or associated with mobile obstacles with specific size and velocity

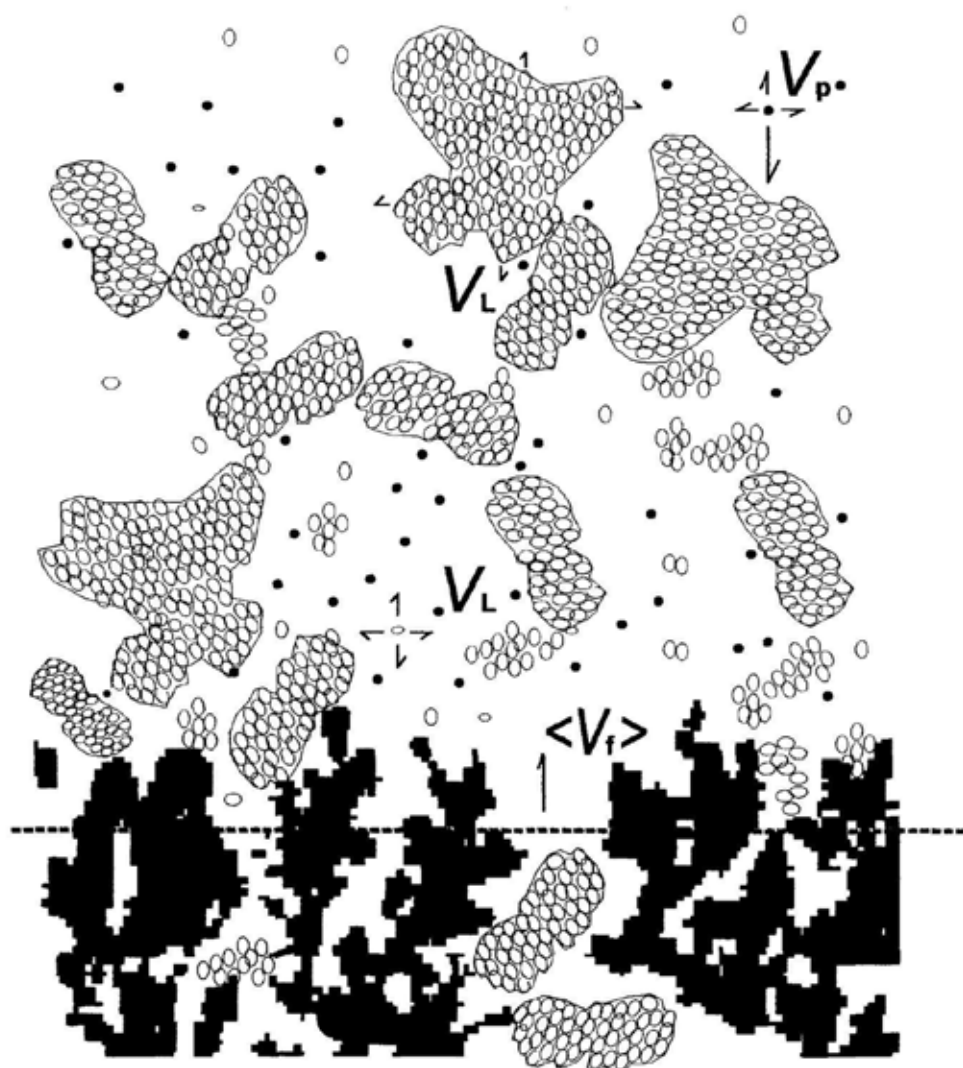


Figure 3. Scheme for the growth of silver electrodeposits in a gelled medium. Large percolated agglomerates and small colloidal particles characterized by wide size and velocity distribution functions. Black dots denote silver ions in the solution; white ellipsoids represent agarose molecules that form a number of free and partially percolated agglomerates; black areas correspond to electrodeposited silver.  $\langle V_f \rangle$  is the average growing front velocity;  $V_p$  is the velocity of electrodepositing ionic species in the solution;  $V_L$  denote the displacement velocity of agarose agglomerates in the medium. The horizontal dashed trace line represents the average plane of the growth front.

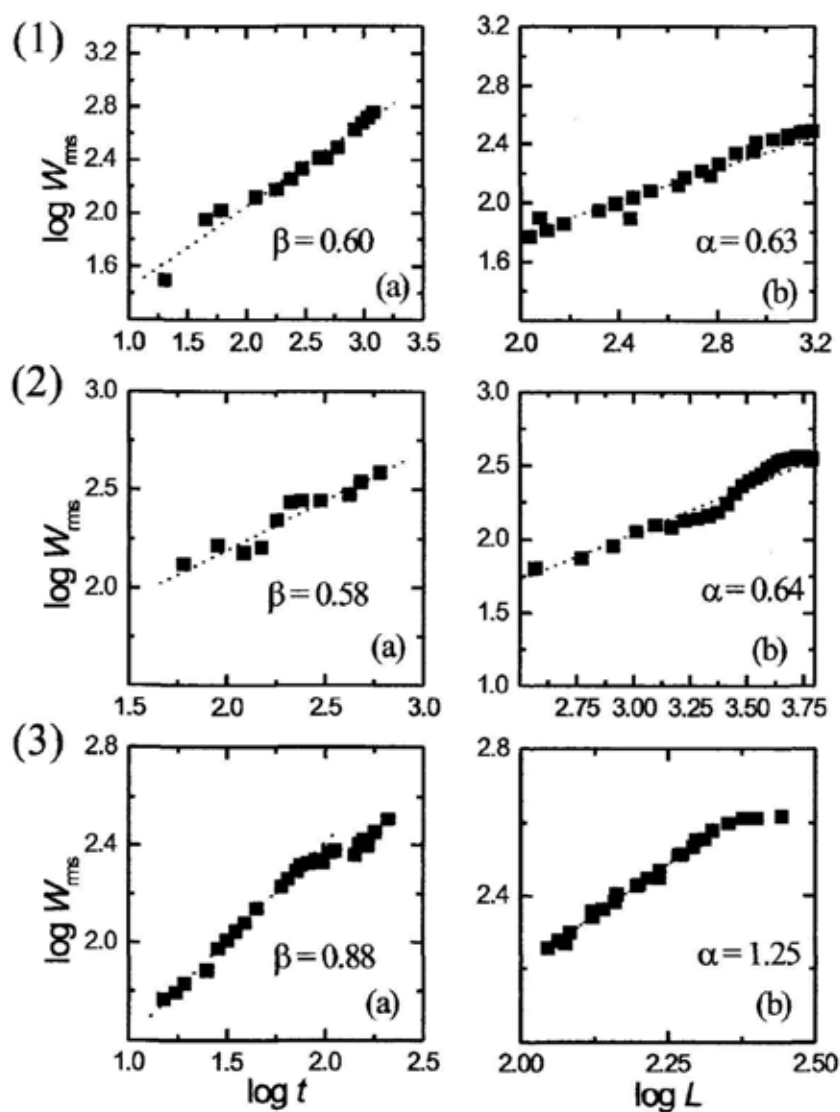


Figure 4.  $\log W_{rms}$  versus  $\log t$  plots (a), and  $\log W_{rms}$  versus  $\log L$  (b) plots, for the AG (1), FP (2), AG (3). Plating bath composition and electrodeposition conditions are the same as indicated for Fig. 2.

distribution functions. These facts allowed us to distinguish between localized and non-localized random field pinning. This distinction provides a possible reason for the anomalous roughness behavior as compared to the KPZ equation for the AG medium.

## Acknowledgments

This work was financially supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) and PICT 98 No. 06-03251 from Agencia Nacional de Promoción Científica y Tecnológica of Argentina.

## References

1. S. J. Han, D. P. Belanger, W. Kleemann and V. Nowak, Phys. Rev. B **45**, 9728 (1992).
2. M. Lederman, J. Selinger, R. Bruinsma, J. Hammann and R. Orbach, Phys. Rev. Lett. **68**, 2086. (1992).
3. H. E. Stanley, A.-L. Barabási, *Fractal Concepts in Surface Growth*, Chapters 9-11, Cambridge University Press, N.Y. (1995).
4. P. Tegzes, T. Vicsek and P. Schiffer, Phys. Rev. Lett. **89**, 094301-1 (2002).
5. T. Vicsek, M. Cserzo and V. K. Horvth, Physica A **167**, 315 (1990).
6. R. Lenormand, E. Touboul and C. Zarcone, J. Fluid Mech. **187**, 165 (1988).
7. M. A. Rubio, C. A. Edwards, A. Dougherty, J. P. Gollub, Phys. Rev. Lett. **63**, 1685 (1989).
8. P. Meakin, A. Birovljev, V. Frette, J. Feder and T. Jossang, Physica A **191**, 227 (1992).
9. S. He, G. L. M. K. S. Kahanda and P-Z Wong, Phys. Rev. Lett. **69**, 3731 (1992).
10. S.V. Buldyrev, A.-L. Barabási, F. Caserta, S. Havlin, H. E. Stanley and T. Vicsek, Phys. Rev. A **45**, R8313 (1992).
11. S. V. Buldyrev, A.-L. Barabási, S. Havlin, J. Kertész, H. E. Stanley and H. S. Xenias, Physica A **191**, 220 (1991).
12. J. Zhang, Y.-C. Zhang, P. Alstrom and M. T. Levinsen, Physica A **189**, 383 (1992).
13. J. Kertész and T. Vicsek, J. Phys. A **19**, L257 (1986).
14. R. Bruinsma and G. Acapli, Phys. Rev. Lett. **52**, 1547 (1984).
15. J. Koplik and H. Levine, Phys. Rev. B **32**, 280 (1985).
16. M. Kardar, G. Parisi and Y.-C. Zhang, Phys. Rev. Lett. **56**, 889 (1986).
17. E. Medina, T. Hwa, M. Kardar and Y.-C. Zhang, Phys. Rev. A **39**, 3053 (1989).
18. C.-H. Lam and L. M. Sander, Phys. Rev. Lett. **69**, 3338 (1992).
19. H. Horvth, F. Family and T. Vicsek, J. Phys. A **24**, L 25 (1991).
20. J. G. Amar, P. M. Lam, F. Family, Phys. Rev. A **43**, 3053 (1989).
21. L.-H. Tang, M. Kardar and D. Dhar, Phys. Rev. Lett. **74**, 920 (1995).
22. D. A. Kessler, H. Levine and Y. Tu, Phys. Rev. A **43**, 4551 (1991).
23. W. Kinzel in *Fractals and Disordered Systems*, A. Bunde and S. Havlin, Eds.,

Springer, Heidelberg (1991).

24. L.-H. Tang and H. Leschhorn, Phys. Rev. A **45**, R8309 (1992).
25. H. Leschhorn and L.-H Tang, Phys. Rev. E **49**, 1238 (1994).
26. Z. Olami, I. Procaccia and R. Zeitak, Phys. Rev. E **49**, 1232 (1994).
27. M. A. Pasquale, S. L. Marchiano, A. J. Arvia, J. Electroanal. Chem. **532**, 255 (2002).
28. M. A. Pasquale, S. L. Marchiano, A. J. Arvia, in preparation.
29. M. A. Pasquale, S. L. Marchiano, A. J. Arvia, J. Appl. Electrochem. (in press).
30. F. Family, T. Vicsek, *Dynamics of Fractal Surfaces*, World Scientific, Singapore, (1991).
31. P. L. Schilardi, O. Azzaroni, R. C. Salvarezza, A. J. Arvia, Phys. Rev. B **59**, 4638 (1999).
32. H. Leschhorn, Physica A **195**,324(1993).

This page intentionally left blank

# EPIDERMAL RIDGES : POSITIONAL INFORMATION CODED IN AN ORIENTATIONAL FIELD

MINH BINH NGUYEN, VINCENT FLEURY AND JEAN-FRANÇOIS GOUYET

*Laboratoire de Physique de la Matière Condensée, Ecole Polytechnique 91128 Palaiseau cedex France*

*E-mail: vincent.fleury@polytechnique.fr*

The study of the emergence of structures in biological systems, complex in essence, is of a great interest. It has been observed for a long time that the shape of fingers is correlated to the "drawings" of the epidermal ridges. But the exact relationship between the final 3D shape of the finger and the 2D epidermal pattern remains unclear. We show here that the pattern of the epidermal ridges explains naturally several features of the final shape of fingers, after the pattern is established.

## 1 Introduction

The main purpose of this paper is to show that an orientational field plays a crucial role in the formation of biological structures, frequently branched like in lungs, kidneys, and even in plants. How does an orientational order influence morphogenesis? This question may be asked in presence of many biological structures, starting at a small scale by sub-cellular structures such as the mitotic spindle, and going up to individual plant cells such as root hairs, and to more macroscopic structures, like trees, vegetables or animal organs. All parts of animals and plants are formed with fibers and cells which, in most cases if not all, exhibit an orientational order. This orientational order is certainly linked to both the elongated nature of biopolymers, and to the anisotropy of individual cells, such as fibroblasts<sup>1,2</sup>. In most cases, these two features are not independent. In the case of animal tissue, the biopolymers are essentially the collagens, the keratin, the chitin. Depending on the tissue, up to 80% of the dry weight of mammalian organs, such as the lung, or even of skin, may be collagen. In animal tissue, the orientational order is often very conspicuous: abdomen of insects, cartilage rings of the lungs, stacks of rings of arthropod antennae, etc. all show regular orientational order, in the form of rings going around an ellipsoidal or a cylindrical structure<sup>3,4,5</sup>. Now, one should notice that regular shapes, such as fruits, or antennae of arthropods may be fibered very simply, while more complex branching shapes will be fibered in a very complex pattern which spouses the branching structure intimately (figures 1 and 2<sup>6,7</sup>). Since the tissue is generally intrinsically fibered, one should wonder how the fibered nature of the tissue contributed to the branching morphogenesis. We shall not treat here the case of a branching structure, but concentrate, as a first step, on simple spheroidal shapes.

However, even for simple fibered spheroidal "shells" there exist complex orientational situations, in the form of loops, whorls, or arches (figure 3), these are the classical names, in forensic science of the dermatoglyphs, *i.e.*, the complex drawings made by the epidermal ridges (the "fingerprint" is the ink or grease print left on an object, *e.g.* paper). Another very frequent structure is the triradii (see figure 3b). These topologies are actually classically described in physics of liquid crystals by a



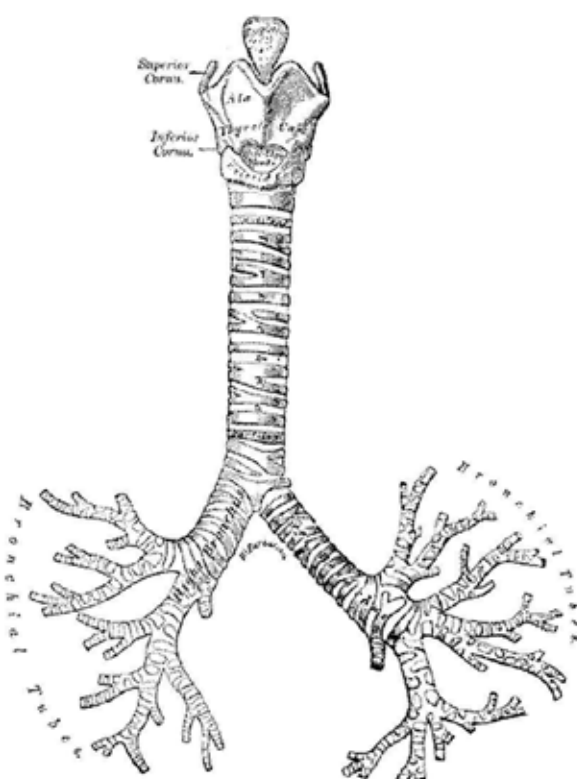


Figure 1. The fractal branching complexity in the lung: the fibered structure of the mesenchyme plays a crucial role in the morphogenesis of organs. In this classical engraving (Gray's anatomy book) one sees the cartilage rings and triradii at bifurcations, which play a role in the mechanical equilibrium of the first branching points.

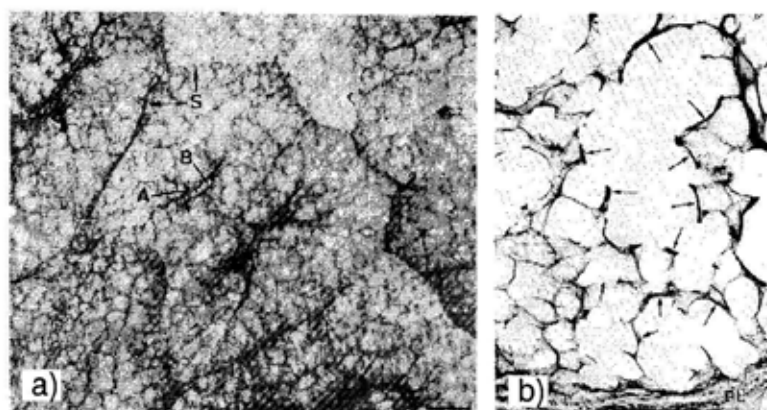


Figure 2. In more distal parts of the lung, the alveolae (right) are all built with collagen fibers stretched around alveolae. The lung structure is akin to a very fibered branching "foam" (left). The collagen is laid down by fibroblasts during growth of the branches.

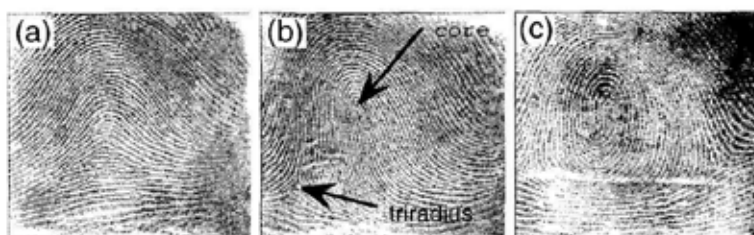


Figure 3. Three types of fingerprint patterns: arch (a), loop (b), whorl (c). Arches dress 5% of fingers, loops 60% and whorls 30%.

singularity index. They correspond to topological singularities of the orientational field. The triradii, for example, is, at least topologically, the  $-1/2$  singularity of the theory of liquid crystals<sup>8,9</sup>. In this article, we wish to address the question of the relationship between the “drawing”, or pattern, of a 2D orientational field, which dresses tangentially a 3D structure, and the overall morphology of the said 3D structure. The simplest instance of such a situation is the shape of the soft tissue of the last phalanx of fingers. Other situations may be found in fruit or vegetable shapes. We will show that a presumptive spherical shell bearing an orientational field akin to liquid crystals is deformed with respect to sphericity. A bump should be expected on the shell, in the region of the singularities. However, contrary to intuition, the bump is not located in the center of the singularity, but is shifted to another position, as will appear clear in the sequel. Prior to going into the technical details of the model, we wish to recall a few facts about epidermal ridges.

## 2 Epidermal Ridges

### 2.1 General facts

Terrestrial vertebrates exhibit a bare skin surface at the tips of fingers. The skin in this area, especially on toes, is much thicker than the skin in other places of the body. There exists micro-reliefs in the millimeter or sub-millimeter range which may be in the shape of dots (pegs, warts), in the shape of lines (epidermal ridges -ER) or in the shape of aligned dots. Epidermal ridges are found on human fingers, on many primates fingers, and even on fingers of animals which are more remote phylogenetically, especially the koalas. Epidermal ridges are also found at the end of the tail of many monkeys. Ridges with larger dimensions are found also on the skin of many animals, often on the face. Considering the epidermal ridges of humans, it must be said that there exist a few genetic conditions such as absence of ER<sup>10</sup> (extremely rare), or aligned dots instead of lines<sup>10</sup>. In the case of absence of ER, the skin is often “bad”, sweat glands are insufficient, the skin is dry and painful, and it exhibits cracks and crevasses<sup>11</sup>.

## 2.2 *What are epidermal ridges?*

Ridges appear as "folds" of the skin. They are found on the palmar side of the fingers and on the hand, although they are not as pronounced in the region of the wrist. In some mutants, one observes very small nails, and sometimes even an absence of nail, in those cases, the epidermal ridges may well go all around the phalanx<sup>12,13</sup>. The "folds" are approximately as thick as 3 times their width. This is to say that the pattern is quite deeply rooted, down to the dermis<sup>14</sup>. This is apparent in the fact that they are very difficult to remove or alter, and, unless the skin is deeply damaged, they will regenerate. The "folds" are made of several layers, which do not have strictly the same wavelength. The surface ridge breadth encompasses actually two wavelengths of the deeper primary ridge system, found in the stratum basalis (deeper skin). At time of formation, the surface ridge breadth is of the order of 50 micrometers, and grows linearly with embryo size<sup>14</sup>. Sweat glands open up in the bottom of the furrows of the folds of the stratum basalis, but they open up on the surface on the summit of the folds (which have a doubled wavelength). Cells are smaller and more numerous in the bottom of the crevasses, in the deep layers. The uppermost layers are composed of dead cells which chip away on top as the skin regenerates in the bottom. The skin tissue is a composite material, which also comprises vessels and nerves. There exist correlations between the position of nerves, vessels, and the structure of the interstitial tissue, but we shall consider the skin as some uniform medium.

## 2.3 *Genetics of epidermal ridges*

Epidermal ridges are not strictly predetermined genetically. The correlation of patterns between identical twins is statistically significant, but only slightly better than between left and right hands of an individual<sup>11,15</sup>. This is to say that the patterns may sometimes be very close, but it is common that they can be completely different. There are epigenetic factors. There exist statistically significant relationships between many medical disorders and the shape of ER<sup>11,10</sup>. It seems that the patterns are attractors of some process, which can give different outcomes, although it is strongly influenced by genetics. A reasonable hypothesis is that general features such as length and width of the fingers are under genetic control, while the detailed drawing is more variable and prone to depend on epigenetic cues. A proof that epidermal ridges are not themselves genetically coded is found in aberrant fingers. Very severely malformed fingers, especially branching fingers, have most generally well ordered ER which dress the finger in a pattern that adapts itself to the malformation<sup>13</sup>. It sounds very unlikely that a set of genes would have been ascribed to making a specific ER for a very malformed finger. Also, it seems that mechanical properties of the tissue, which play a role in the problem (see below) are affected by mutations concerning collagen anlagen, or osmotic balances. Hence, a natural correlation exists between the formation of ER and possibly other diseases or physical characters.

## 2.4 When and how do they form?

Epidermal ridges form between the 8th and 15th week of gestation. The formation of the epidermal ridges starts distally, and proceeds towards the more proximal part of the finger. The appearance of ER is correlated to the presence of mount elevations on the fingers and hands called the volar pads. There are 11 possible locations of such mount elevations on the finger and hand<sup>12</sup>. They truly "correspond" to the position of the whorls, arches, loops, and triradii which anyone may find on his hand, although the correlation may be inverted. For example, close to the fourth finger, in the region where the proximal phalanx joins the hand there is generally a bump, in the adult, which was actually a dip, in the embryo, between two pads. During growth, it seems that the pattern was somewhat buckled upwards, but keeping the correlation with the initial structure. The mount elevations named volar pads exist in some animals in the adult, in the form of walking pads, as in cats. However, in humans, the mount elevations appear only transiently as swellings, and they disappear progressively during the period of formation of the ER ("involution" of the volar pads). Therefore, it is considered in prenatal anatomy that ER appear in response to stresses in the volar pads, as they regress, contract or shrink. There is no dynamic study of this phenomenon which is not surprising, considering that it occurs in utero, on an embryo about 2 months old. After regression of the volar pads, and formation of the ER, there is a constant growth of the fingers, with an increase in tension of the skin<sup>15,14</sup>.

The question of the exact mechanism of formation of the ridges is not yet answered, although it may be speculated that mechanical factors orient cell division in a way that generates "folds". Indeed, classical biological studies have shown, apparently, that the skin does not actually fold<sup>12,14</sup> (hence the comas above). According to biologists "the epidermal ridges are not folds resulting from movement of the parts but are rather the result of epidermal proliferation"<sup>14</sup> (p.161, and references therein). It is a differential development of cell layers which generate the apparently "folded" structure. We shall now on use the word folds without comas, although it is understood that they appear by differential growth. This description is reminiscent of the Grynfeld instability<sup>16</sup> in physics by which folds form, not by buckling, but by surface diffusion which reorganizes the surface in an apparently buckled shape, in order to accomodate the elastic energy imposed by a uniaxial (tangent to the surface) stress. Such differential buckling is also observed in brain folding<sup>4</sup>. In this case, there is some information about the individual motion of cells: cells divide actively in the bottom of the fjords and migrate to the top of the folds, with a motion which is perpendicular to the fold direction (they climb uphill the giry). This is compatible with what is observed in skin<sup>14</sup>. On more fundamental grounds, study of individual cells have shown that cells tend to divide more actively in response to tension. From a physical point of view, it is well known that stresses are higher in the bottom of the furrows<sup>16</sup>. It makes sense that the bottom of the furrows should lead the cell division, and be the pool of new cells. A significant observation in, this context is that the relative thicknesses of the cell layers change during formation of the ER; it has long been suggested that this produces stresses which stimulate growth<sup>14</sup>. Another significant observation

is that, in the late stage of formation of ER (115mm-140mm), new ridges appear not by cellular proliferation, but at the expense of existing ridges. The skin is at that moment stretched and, as the ridges get away, new ridges appear in between by reorganization of the former ridges. This is even more reminiscent of a Grynfeld instability. Indeed, in this case, the tissue seems to reorganize itself to reduce uniaxial tensile stress (such an instability works with tensile as well as compressive stress). For a detailed description of the anatomy of the cell layers during formation of ER, the work of Hale <sup>14</sup> is of considerable interest.

Now, in anatomy, several global morphological observations suggest a role of mechanical factors. Indeed, the shape of the fingers is correlated to the drawing of the ER. Especially, more elongated and narrow fingers are significantly correlated to the presence of loops <sup>15,11</sup>. Penrose <sup>17</sup> considers that the lines of the ER spouse the lines of principle curvature of the shape of fingers (especially the pads) at time of formation. This idea is essentially based on the observation that in many abnormal fingers, especially in the absence of nails, the ER simply makes stacks of rings with the axis oriented in the direction of the finger, and with circles running around the finger (this situation is the one generally observed on the middle phalanx, although the circles are interrupted on the dorsal side of the fingers) <sup>13</sup>. This idea, although suggestive of an anisotropy in the problem, does not clarify the mechanism of formation. If there is an anisotropic differential growth in reaction to tension, then one would expect the skin to extend and "buckle" by cellular proliferation in the other direction, and form folds parallel to the finger axis (the tension being higher azimuthally because of a smaller curvature). An argument contrary to this one would be that the finger, and especially the cartilages, extend by growth and tend to stretch the whole finger. Although a static view would consider that the tension is higher in the azimuthal direction, the force is greater in the direction of growth of the finger. This is apparent in the fact that the finger extends forward more than radially. Then, a possibility arises that the skin over-responds to such a stretch by a proliferation which is not in pace with the growth. If the skin extends more than the finger, it has to buckle in order to accomodate the extra-matter (an opposite effect is observed in shar-peï dogs, for example, which are very folded at birth; the folds of the baby shar-peï progressively decrease as the body of the dog gets bigger, because the skin, after birth, does not extend as rapidly as the body of the animal).

In such a view, the extension is by tension, and not truly by curvature. Another possibility is that tension in one direction induces proliferation in the perpendicular direction. Another feature which should be taken into account is the intrinsic anisotropy of the mechanical properties of such a folded material. Indeed, the bending modulus of a folded material will be lower in the direction perpendicular to the folds than in the direction parallel to the folds <sup>5</sup>.

Still, these models all suppose that formation of the folds, and especially the final 2D pattern, follows passively the dynamics and morphology of the finger growth, including of the volar pads, and neglects completely the fact that the folds themselves have their own dynamics and that they will induce morphological consequences on the growth of fingers. Even if the volar pads are responsible for the establishment of the pattern, the composite material of which the skin is made exhibits, in the end, a very conspicuous orientational order, which has itself morphological consequences.

If, as is observed, the skin tension increases in the last phalanx during growth, the drawing of the ER will play a role in the final finger shape. In the end, it is more than probable that the 3D shape of the finger and the 2D pattern all form hand in hand.

We shall consider in the rest of this article only one aspect of the problem, in that we will assume that a folding mechanism by differential growth is at work which generates the patterns of ER as we see them. We then admit that there is an orientational order with a vector  $\mathbf{t}$  defined everywhere by the local direction of the ER, which is recorded for good in the skin. We will suppose that the ridged layer is a thin shell bearing an internal turgor pressure, as is likely during embryo development. We then address the following question: if the turgor pressure is itself uniform, would the last phalanx of any finger, covered with a pattern of ER, be a cylinder with a hemispherical cap? We will show with a simple model that this is not the case, and that the shape would depart from sphericity. It comes as a surprise that the shape is qualitatively one of a finger. This implies that the orientational order plays an important role in the morphology, and probably the dynamics of growth, of a biological tissue.

### 3 The model

To model the impact of the epidermal ridges on the shape of the palmar side of fingers, we suppose that the elasticity along the ridges, more keratinized, is different from that in the furrows (grooves) between the ridges. This orientational order leads to a stiffness locally higher in the longitudinal directions parallel to the ridges than in the transversal directions. During the growth of the fingers, the local pressure due to the multiplications of dermal cells, leads to form the final "equilibrium" shape of the fingers in a manner somewhat similar to the equilibrium shape of a crystal. It is well known that the shape of crystals depends on how the surface tension varies with orientation. In our case, we have to consider a special kind of crystal: one whose surface tension depends on the drawing of the lines.

In a simple approximation, the ridge-bearing surface of the skin is considered as a membrane, dressed by a vector field of modulus 1 tangent to the surface and with a varying in plane orientation  $\mathbf{t}(\mathbf{r})$  the ridges at point  $\mathbf{r}$ . The distortion energy due to the anisotropy of this vector field contains at order  $(\nabla \mathbf{t})^2$ , three main contributions, associated to three elastic constants. These constants correspond to three basic types of deformation, splay, twist and bend. These deformations relate the tendency of the director (here the direction of a tangent to the fiber structure) to remain parallel to restoring torques throughout the media. The three constants of importance are: splay – a change in the direction of the director when moving at right angles to the director; bend – a change in the direction of the director when traveling in the direction of the director; twist – a change in the direction of the director when moving out of the direction of travel at right angles to the director. The corresponding free energy density, called the Oseen-Frank free energy is,

$$F_d = K_1(\text{div } \mathbf{t})^2 + K_2(\mathbf{t} \cdot \text{rot } \mathbf{t} + \tau)^2 + K_3 |\mathbf{t} \times \text{rot } \mathbf{t}|^2 + (K_2 + K_4)(\text{tr}(\text{grad } \mathbf{t})^2 - (\text{div } \mathbf{t})^2)$$

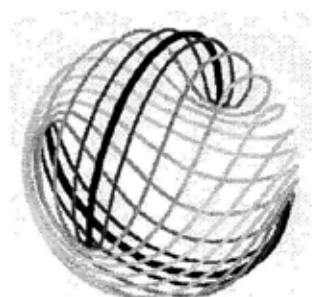


Figure 4. A sphere dressed with our 3D epidermal ridges. The ridges are obtained by solving the Laplace equation with boundary conditions on the two singular lines

The constant  $K_2, i = 1, 2, 3$  correspond to the splay, twist and bend elastic constant respectively. They give the energetic cost of a non parallel configuration in terms of these three pure (independent) distortions. The parameter  $\tau$  represents an intrinsic elastic stress in chiral materials. In the present study we only consider the dominant splay contribution with,

$$F_d = K_1(\text{div } \mathbf{t})^2$$

The Frank's constant  $K_1$  is positive, ensuring that the deformed membrane corresponds to a minimum of energy.

### 3.1 Modeling the ridges

The fingerprints belong mainly to three classes, and among these classes, the fingerprints having a ridge configuration with a loop core (figure 3b), are the most common. They represents 60% of all the cases. We have roughly simulated this pattern using the image of a tennis ball, as shown in figure 4. This makes the calculations simpler.

In figure 4, the black lines are singular and join two cores. The various lines simulating the ridges are then obtained as equipotential curves between the two extreme singular curves. To the upper one the value of the potential is taken equal to +1, to the lower one we have attributed the value -1. So, such a drawing would correspond to the palmar side of a finger, to which we have performed a symmetry with respect to the plane of the nail, and a symmetry with respect to the plane separating the two last phalanx.

From the curves, simulating the ridges, we can extract the free energy contribution

$$F = F_0 + F_d = F_0 + K_1(\text{div } \mathbf{t})^2 \quad (1)$$

Figure 5 shows the resulting Frank elastic energy of the surface.

We observe a divergence of the energy density in the four directions corresponding to the four loop cores. Notice that due to the symmetries only one of them will correspond to the palmar region of a finger.

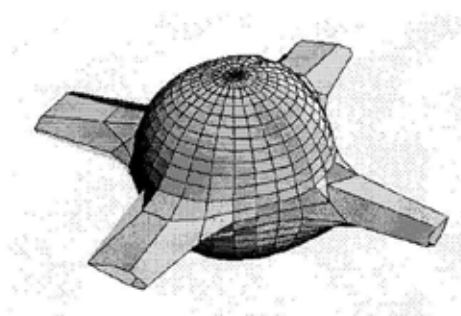


Figure 5. Frank energy associated to the ridges in figure 4. We use spherical coordinates with a radius proportional to  $F$

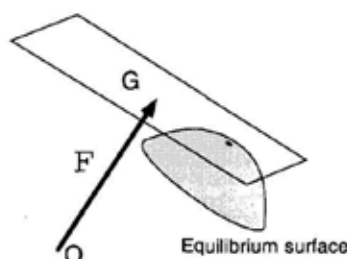


Figure 6. Principle of the Wulff construction.

### 3.2 Constructing the equilibrium shape

The construction of the equilibrium shape from the surface tension of the membrane, can be done using Wulff's construction<sup>18</sup>, as for crystalline shapes. In spherical coordinates, we plot the rays  $OG$  where  $OG = F \mathbf{n}(\mathbf{r})$ , where  $\mathbf{n}(\mathbf{r})$  is the normal vector at a point  $\mathbf{r}$  of the undeformed sphere (figure 5),  $F$  is given by equation (1). The equilibrium shape is obtained as the envelope of the planes normal to  $OG$  at point  $G$ . Let us recall that a tension is associated to an interfacial energy. if the interfacial energy is constant, then the surface is spherical (soap bubble). But, if the interfacial energy varies, so does the tension in the surface, and the calculation of the shape is ore complex. The shape has to balance the distribution of tensions with the pressure exerted perpendicularly to the curved surface at each point. This is what the Wulff's construction provides. The construction is shown in figure 6.

## 4 Results

The result of the effect of the orientational field possessing a core singularity, is shown in figure 7. In principle the process has to be performed iteratively, and this equilibrium shape taken as a new starting point to recalculate the diffusion field and the the Frank elastic energy. But this appears to be a second order contribution.





Figure 7. Equilibrium shape of a membrane constrained by the orientational field defined in figure 5. It is the envelope of the planes defined via a Wulff construction

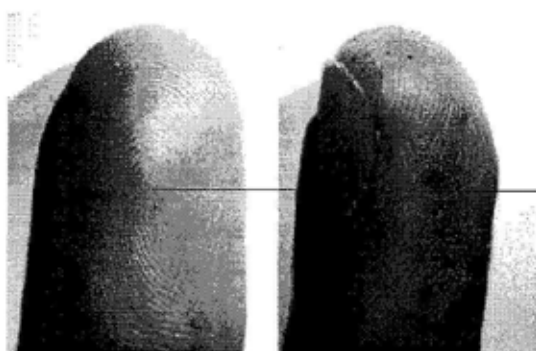


Figure 8. A typical finger with the epidermal ridges showing a (displaced) bump in the region of the core.

We observe that these results also explain the particular position of the bump on the palmar side of the fingers. As clearly visible in figure 8 the summit is not right at the core of the epidermal ridges pattern, but somewhat more distal. Anyone may check this on his or her own fingers.

Let us consider now our singular surface energy (figure 5), it has four singularities, one pointing towards the reader in figure 4 and 5. We observe that the drawing of the lines is not symmetrical around the axis (there is a loop instead of a target, which breaks the symmetry). Let us make a schematic 2D cross-section in a plane passing through the core, and along the singular line of the loop (figure 9, left). We see that although the singularity in surface tension is at the point  $O$ , the cusp in the shape is not. The polar equilibrium shape presents a cusp shifted with respect to the  $Oz$  axis. The envelope from the Wulff construction is shown on the right. It confirms the asymmetry.

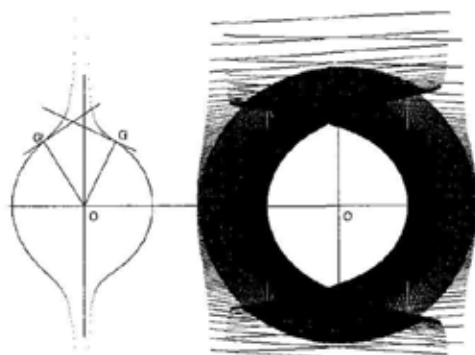


Figure 9. Asymmetry of the bump. The core is on the  $Oz$  axis while the summit of the bump is displaced towards the left

## 5 Conclusion

The protruding feature in the region of the core finds itself (as anatomically observed) somewhat displaced from the position of the core. In other words, the very centre of the "epidermal ridges" of the dressed sphere is not at the summit of the relief. This fact has a more deeply rooted origin. When an equilibrium shape is formed with a symmetrical surface tension across a given direction (say for example a square symmetry), the corresponding shape has the very same symmetry and the same direction of mirror symmetry. However, if an equilibrium shape is formed with an asymmetrical surface tension with respect to a plane or axis, then, not only the resulting shape is not symmetrical around the direction that breaks the symmetry, but, in addition, the very position of the direction of asymmetry is shifted towards the "softer" region. This is clearly shown in figure 9, on a 2D calculation, in polar coordinates, which would be the analog of a cut across the last phalanx, by a vertical plane passing through the nail and the core, down to the first joint. This gives a simple explanation to the mismatch between the position of the core of ER (symmetry breaking of the orientational field), and of the actual position of the summit of the finger pad (symmetry breaking of the corresponding shape). The more general consequence of this fact is that the local orientational field  $\mathbf{t}(x, y, z)$  which is imparted on skin cells contains a great deal of morphological information and hence part of the "positional information" leading to the exact finger shape<sup>19</sup>. Therefore, the fields of "morphogens" are not solely the concentrations of all possible bio-chemicals in the problem, but also a vector, "recorded" in the orientational field. This fact, of course, may have much wider consequences. The energetics of an orientational field implies specific morphological tendencies for given topological configurations, a fact well admitted in physics of liquid-crystals. In the biological case, the position of holes, dimples or protrusions, the orientation of defenses, pins, darts, roots or branches and the speed of growth of such organs, will all be influenced by the presence of orientational singularities. We acknowledge

that the bio-mechanics of a growing tissue is certainly more complex in the general case, than the model which we have used.

## References

1. S. Gilbert, *Developmental Biology*, Sinauers (4th edition), Sunderland Mass., (1994).
2. J. Bard, *Morphogenesis*, Cambridge University Press, Cambridge (1992) Chapter 5.
3. Y. Bouligand and F. Livolant, The organization of cholesteric spherulites, *J. Physique*, **45**, 1899 (1984).
4. V. Fleury, *Des pieds et des mains*, Flammarion, Paris, (2003).
5. V. Fleury and T. Watanabe, *C. R. Acad. S. Biologies*, **325**, 571 (2002).
6. H. Gray, *Anatomy of the human body*, Lear & Febiger, Philadelphia, (1918).
7. E. Weibel, *The pathway for oxygen. Structure and function in the mammalian respiratory system*, Harvard University Press, Cambridge Mass. (1984).
8. P. G. De Gennes and A. Prost, *The physics of liquid crystals*, Clarendon, Oxford (1993).
9. P. Pieranski and P. Oswald, *Les cristaux liquides*, Gordon and Breach (2000).
10. B. Schaumann and M. Alter, *Dermatoglyphics in Medical Disorders*, Springer Verlag, New York Heidelberg Berlin (1976).
11. N. M. Durham, *Trends in dermatoglyphic research*, Kluwer Academic Publishers, Dordrecht, Boston, London (1990).
12. H. Cummins, The topographic history of the volar pads (Walking pads, tast-ballen) in the human embryo, *Contributions to Embryology* N°13, 103-126.
13. H. Cummins, *Amer. Jur. Anat.*, **38**, 89 (1926).
14. A. R. Hale, *Am. J. Anat.* **91**, 147 (1952).
15. P. T. Babler, *Coll. Antropol.* **11**, 297 (1987).
16. J. Müller, *Study of stress induced morphological instabilities*, PhD Thesis, Department of Physics, Mc Gill University, Montreal, (1998).
17. L. S. Penrose, *Nature*, **205**, 544 (1965).
18. C. Godrèche, *Solids far from equilibrium*, Aléa Saclay (1998).
19. L. Wolpert, Positional information and pattern formation, *Curr. Top. Dev. Biol.* **6**, 183-224 (1971).

# MULTISCALE PRINCIPAL COMPONENTS

ANTOINE SAUCIER

*Mathématiques appliquées et génie industriel  
École Polytechnique de Montréal  
Montréal (Québec), Canada, H3C-3A7  
E-mail: Antoine.Saucier@polymtl.ca*

We show that data-adaptive orthogonal wavelet bases can be obtained from a generalization of principal component analysis. These bases are not self-similar in general. They are optimal for data-compression and denoising. **Keywords:** Data-adaptive wavelets, orthogonal basis, principal component analysis

## 1 Introduction

### 1.1 Wavelets and principal component analysis

Since the introduction of wavelets, we have seen a proliferation of different wavelet bases. The choice of an intrinsically well-adapted wavelet for the analysis of a given class of signals is therefore a non trivial task. In this paper, we propose a new approach for the construction of data-adaptive orthogonal wavelet bases. Our approach is based on a generalization of principal component analysis.

Most applications of wavelet bases exploit their ability to efficiently approximate particular classes of functions with few non-zero wavelet coefficients (Mallat <sup>6-b</sup>). These coefficients are scalar products  $\psi^T S$  of a wavelet  $\psi$  with the signal  $S$ . If  $S$  is regular and  $\psi$  has enough vanishing moments, then the wavelet coefficients are small at fine scales. Several methods (e.g. Geronimo *et al.* <sup>2</sup>) have been developed to control the magnitude of wavelet coefficients via vanishing moments. Other methods include the matching pursuit algorithm (Mallat and Zhang <sup>5</sup>), the spectral approach of Lilly and Park <sup>4</sup> and wavelet packets (Learned and Willsky <sup>3</sup>). Yiou *et al.* <sup>7</sup> constructed data-adaptive wavelets with principal components (PC). However, these wavelets all have the *same support diameter*, i.e. they do not form a multiscale basis.

None of these approaches aims at the direct minimization of the mean square average wavelet coefficients  $E((\psi^T S)^2)$  for a given reference signal ( $E(\dots)$  denotes an expectation value). In this paper, we propose to adapt principal component analysis (PCA) to the construction of bases that have multiscale compact supports. To our knowledge, this approach to the wavelet adaptivity problem has not been investigated. To connect wavelets with principal components, we construct localized functions that minimize (or maximize) their correlation with the signal of interest, while remaining mutually orthogonal. They have compact supports with variable diameters. These functions are a multiscale generalization of principal components, and therefore we call them *multiscale principal components* (MPCs).

### 1.2 Notations and terminology

We focus on one dimensional real signals defined at discrete coordinates  $x_i$ ,  $i = 1, 2, \dots, N$ . The value of a signal  $f$  at  $x = x_i$  is denoted by  $f(i)$ . The signal is regarded as a column vector  $\mathbf{f} = (f(1), f(2), \dots, f(N))^T$ . All vectors are assumed to be  $N$ -dimensional, unless specified otherwise. The diameter of the support of a function is often called the *size* or *scale* of the function.

## 2 Overview of our approach to the construction of multiscale principal components

We consider the construction of a basis of MPCs denoted by  $\phi_1, \phi_2, \dots, \phi_N$ . We assume that the  $\phi_n$ s are *localized functions*. In this paper, we say that a function is localized if its support is of the form  $[n, m]$ , where  $n$  and  $m$  are finite integers. The support diameters of the  $\phi_n$ s, denoted by  $N_n$ ,  $n = 1, 2, \dots, N$ , are assumed to satisfy  $N_1 \leq N_2 \leq \dots \leq N_N$ . We build the  $\phi_n$ s from *small to large scales*. In the first step, we build  $\phi_1$ , which is the *first order MPC*.  $\phi_1$  is localized on its support  $\mathcal{I}_1$  and is constrained to have unit norm. Under this constraint, it must minimize  $E((\phi_1^T \mathbf{F})^2)$ . In the second step, we build  $\phi_2$ , which is the *second order MPC*.  $\phi_2$  is localized on its support  $\mathcal{I}_2$ , has unit norm and must be orthogonal to  $\phi_1$ . Under these two constraints,  $\phi_2$  must minimize  $E((\phi_2^T \mathbf{F})^2)$ . More generally, at the  $k^{\text{th}}$  step, we build  $\phi_k$ , which is the MPC of order  $k$ .  $\phi_k$  is localized on its support  $\mathcal{I}_k$ , has unit norm and must be orthogonal to  $\phi_1, \phi_2, \dots$  and  $\phi_{k-1}$ . Under all these constraints,  $\phi_k$  must minimize  $E((\phi_k^T \mathbf{F})^2)$ . We iterate this process until a complete basis is obtained. The final result is an orthonormal basis of MPCs for which the energy contained in the small scale coefficients has been minimized. Consequently, most of the energy is contained in the large scale MPCs.

## 3 Construction of the first order MPC

### 3.1 The optimization problem

We want to construct a normalized function  $\phi_1$  that has minimum correlation with a given reference signal, but which has a compact support  $\mathcal{I}_1 = [k_1, k_1 + N_1 - 1] \subseteq [1, N]$  of width  $N_1$ . The reference signal  $\mathbf{F} = (F(1), F(2), \dots, F(N))^T$  is assumed to be a stationary random process. We define  $\phi_1$ , the first order MPC, to be the solution of the optimization problem

$$\begin{cases} E((\phi_1^T \mathbf{F})^2) \text{ is minimum} \\ \phi_1^T \phi_1 = 1 \end{cases} \quad (1)$$

(1)-a expresses the minimum correlation criterion, whereas (1)-b is a normalization constraint.  $\mathbf{F}$  being stationary, the non-zero components of  $\phi_1$  will be independent of the location of  $\mathcal{I}_1$  within the range  $[1, N]$ . For simplicity, we will therefore solve (1) with  $\mathcal{I}_1 = [1, N_1]$ . Using  $\phi_1^T \mathbf{F} = \sum_{i=1}^{N_1} F(i) \phi_1(i)$ , it follows that  $E((\phi_1^T \mathbf{F})^2)$

can be simplified according to:

$$E((\phi_1^T F)^2) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} C_{i,j} \phi_1(i) \phi_1(j) \quad (2)$$

where  $C_{i,j} \equiv E(F(i)F(j))$  is the *correlation matrix* of  $F$ . We define a vector  $\varphi_1$  of dimension  $N_1$  that contains the first  $N_1$  non-zero components of  $\phi_1$ , i.e.

$$\varphi_1(n) = \phi_1(n), \forall n \in [1, N_1] \quad (3)$$

and a matrix  $C_1$ , of dimension  $N_1 \times N_1$ , by

$$C_1(n, m) = C(n, m), \forall (n, m) \in [1, N_1] \times [1, N_1] \quad (4)$$

Using (3) and (4), (2) becomes  $E((\phi_1^T F)^2) = \varphi_1^T C_1 \varphi_1$  and consequently the problem (1) can be written in the equivalent form

$$\begin{cases} \mathcal{E} = \varphi_1^T C_1 \varphi_1 \text{ is minimum} \\ \varphi_1^T \varphi_1 = 1 \end{cases} \quad (5)$$

(5) is a well-known constrained optimization problem that is encountered in the derivation of principal components (except that we *minimize* instead of maximizing). This problem can be approached with the Lagrange multipliers method. We form the auxiliary function  $U \equiv \varphi_1^T C_1 \varphi_1 - \lambda (\varphi_1^T \varphi_1 - 1)$ , where  $\lambda$  is a Lagrange multiplier. Setting  $\partial U / \partial \varphi_1(i) = 0$  for all  $i$  yields

$$C_1 \varphi_1 = \lambda \varphi_1 \quad (6)$$

which implies that  $\varphi_1$  is an eigenvector of  $C_1$ . Let us denote by  $u_n$  the  $N_1$  normalized eigenvectors of  $C_1$ , and by  $\lambda_n$  the corresponding eigenvalues sorted in decreasing order ( $n = 1, 2, \dots, N_1$ ). With  $\varphi_1 = u_n$ , the function  $\mathcal{E}$  in (5) takes the form  $\mathcal{E} = u_n^T C_1 u_n = u_n^T \lambda_n u_n = \lambda_n$ .  $\mathcal{E}$  is therefore minimum for the eigenvector that has the smallest eigenvalue, i.e.  $\varphi_1 = u_{N_1}$ . We should stress that the first order MPC is not really a new concept. Indeed, it is simply the principal component with minimum eigenvalue for a signal  $F$  which is restricted to the interval  $\mathcal{I}_1$ .

## 4 Construction of higher order MPCs

### 4.1 The optimization problem

Once the first order MPC is obtained, the next step is to construct a second one. More generally, we want to construct the  $k^{\text{th}}$  MPC when the first  $k-1$  MPCs are given. Our goal is to obtain a complete orthonormal basis. We want  $\phi_k$  to be a normalized function that has minimum correlation with  $F$ , that has a compact support  $\mathcal{I}_k = [1, N_k]$ , and that is orthogonal to all the  $N$ -dimensional vectors  $\phi_1, \phi_2, \dots, \phi_{k-1}$ . We will assume that the *supports* are *embedded in each other* according to  $\mathcal{I}_k \supseteq \mathcal{I}_{k-1} \supseteq \dots \supseteq \mathcal{I}_1$ , and that  $N \geq N_k \geq N_{k-1} \geq \dots \geq N_1$ . This

optimization problem can be formulated as follows:

$$\begin{cases} E((\phi_k^T F)^2) \text{ is minimum} \\ \phi_k^T \phi_k = 1 \\ \phi_k^T \phi_{k-1} = 0 \\ \phi_k^T \phi_{k-2} = 0 \\ \dots = 0 \\ \phi_k^T \phi_1 = 0 \end{cases} \quad (7)$$

As in section 3.1, we define the  $N_k$ -dimensional vectors  $\varphi_j$  by

$$\varphi_j(n) = \phi_j(n), \quad \forall n \in [1, N_k], \quad j = 1, 2, \dots, k-1 \quad (8)$$

and the matrix  $C_k$  of dimension  $N_k \times N_k$  by

$$C_k(n, m) = C(n, m), \quad \forall (n, m) \in [1, N_k] \times [1, N_k] \quad (9)$$

The definitions (8) and (9) allow us to reduce the dimension of the problem, which also reduces the computational burden. Using (8) and (9), we obtain

$$E((\phi_k^T F)^2) = \varphi_k^T C_k \varphi_k \quad (10)$$

Using (8), (9) and (10), the problem (7) takes the equivalent form

$$\begin{cases} \varphi_k^T C_k \varphi_k \text{ is minimum} \\ \varphi_k^T \varphi_k = 1 \\ \varphi_k^T \varphi_{k-1} = 0 \\ \dots \\ \varphi_k^T \varphi_1 = 0 \end{cases} \quad (11)$$

The optimization problem (11), which is more complex than the problem (5), can be approached as follows. The orthogonality constraint to the vectors  $\varphi_1, \varphi_2, \dots, \varphi_{k-1}$  defines a *vector subspace* of dimension  $D_k = N_k - (k-1)$ . Let us denote by  $\{\mathbf{P}_i, i = 1, 2, \dots, D_k\}$  an arbitrary orthonormal basis of this subspace. Each  $\mathbf{P}_i$  is orthogonal to all the  $\varphi_n$ s, i.e.

$$\mathbf{P}_i^T \varphi_n = 0, \quad n = 1, 2, \dots, k-1 \quad (12)$$

for  $i = 1, 2, \dots, D_k$ . Such a basis can always be constructed by applying Gram-Schmidt orthogonalization to a collection of  $D_k$  vectors that have been made orthogonal to all  $\varphi_n$ s. In the basis  $\{\mathbf{P}_i\}$ , we denote the coordinates of  $\varphi_k$  by  $y_i$ ,  $i = 1, 2, \dots, D_k$ , i.e.  $\varphi_k = \sum_{i=1}^{D_k} y_i \mathbf{P}_i$ . If  $\mathbf{Y} = (y_1, y_2, \dots, y_{D_k})^T$  and  $\mathbf{P}$  is a (non square) matrix formed of the column vectors  $\mathbf{P}_i$  ( $\mathbf{P}_{i,j}$  denotes the  $j^{\text{th}}$  component of  $\mathbf{P}_i$ ), then it can be shown that  $\varphi_k$  and  $\mathbf{Y}$  are related by

$$\varphi_k = \mathbf{P}^T \mathbf{Y} \quad (13)$$

which is a classical formula from linear algebra (i.e. transformation of coordinates corresponding to a change of basis). Using (13), (11) takes the *equivalent form*

$$\begin{cases} (\mathbf{P}^T \mathbf{Y})^T C_k \mathbf{P}^T \mathbf{Y} \text{ is minimum} \\ (\mathbf{P}^T \mathbf{Y})^T \mathbf{P}^T \mathbf{Y} = 1 \end{cases} \quad (14)$$

On one hand,  $(P^T Y)^T C_k P^T Y = Y^T P C_k P^T Y$ . On the other hand,  $(P^T Y)^T P^T Y = Y^T P P^T Y = Y^T Y$ . Indeed,  $P P^T = I$  because  $P$  is *unitary*, i.e. formed of orthonormal vectors. It follows that (14) reduces to

$$\begin{cases} Y^T (P C_k P^T) Y \text{ is minimum} \\ Y^T Y = 1 \end{cases} \quad (15)$$

The transformation (13) has therefore reduced the optimization problem (11) to the classical principal components problem (15), which is formally identical to (5). The solution of (15) is straightforward:  $Y$  is the eigenvector of  $P C_k P^T$  having the smallest eigenvalue, and then we use (13) to obtain  $\varphi_k$ . It is emphasized that  $P C_k P^T$  is the expression of the matrix  $C_k$  in the subspace of the vectors orthogonal to  $\varphi_1, \varphi_2, \dots, \varphi_{k-1}$ . In that sense, MPCs are truly a multiscale generalization of classical principal components. The procedure described above can be used to construct MPCs iteratively. It can be iterated until a full basis is obtained. In the next example, we will see that data-adaptive wavelets can be constructed by applying this procedure only once.

## 5 Diadic MPCs

### 5.1 Definition

We consider signals of dimension  $N = 2^{n_0-1} L_0$ , where  $n_0 > 0$  and  $L_0 \geq 2$  are integers. Diadic MPCs will be denoted by  $\psi_{n,m}$ , where  $n$  and  $m$  are width and location indices respectively. The  $\psi_{n,m}$ s sharing the same  $n$  have identical sizes  $\ell_n$ , defined by  $\ell_n = 2^{n-1} L_0$  for  $n = 1, 2, \dots, n_0$  and  $\ell_n = N$  for  $n_0 < n \leq n_{\max}$ .  $L_0$  and  $\ell_{n_0}$  are the minimum and maximum function sizes, respectively. If  $n < n_0$ , then the supports of  $\psi_{n,m}$  and  $\psi_{n,m+1}$  are adjacent and disjoint, which implies the orthogonality property  $\psi_{n,m_1}^T \psi_{n,m_2} = \delta_{m_1,m_2}$ . Moreover,  $\psi_{n,m}$  is obtained from  $\psi_{n,1}$  by a simple translation of the support of  $\psi_{n,1}$ . It follows that we need only to define the construction method of  $\psi_{n,1}$  for each  $n$ . If  $n \geq n_0$ , then there is a unique large scale MPC for each  $n$ . The  $\psi_{n,m}$ s form an orthonormal basis of  $\mathcal{R}^N$ , so that any signal  $F$  can be expanded as

$$F = \sum_{n=1}^{n_{\max}} \sum_{m=1}^{M(n)} (F^T \psi_{n,m}) \psi_{n,m} \quad (16)$$

where  $M(n)$  is the number of MPCs of size  $\ell_n$ , given by  $M(n) = 2^{n_0-n}$  for  $n = 1, 2, \dots, n_0$ , and by  $M(n) = 1$  for  $n_0 \leq n \leq n_{\max}$ . The location of the supports of the MPCs is illustrated in figure 1.

### 5.2 Construction

The first step is to build the MPCs of *first generation*. They have equal width  $\ell_1$  and adjacent but disjoint supports. They are denoted by  $\psi_{1,k}$ ,  $k = 1, 2, \dots, M(1)$ , and their supports are  $[(k-1)\ell_1 + 1, k\ell_1]$ , where  $1 \leq k \leq M(1)$ . We build  $\psi_{1,1}$  by solving the optimization problem (1):  $\psi_{1,1}$  is normalized and minimizes  $E((F^T \psi_{1,1})^2)$ . There is only one normalization constraint to satisfy, hence  $\psi_{1,1}$  is



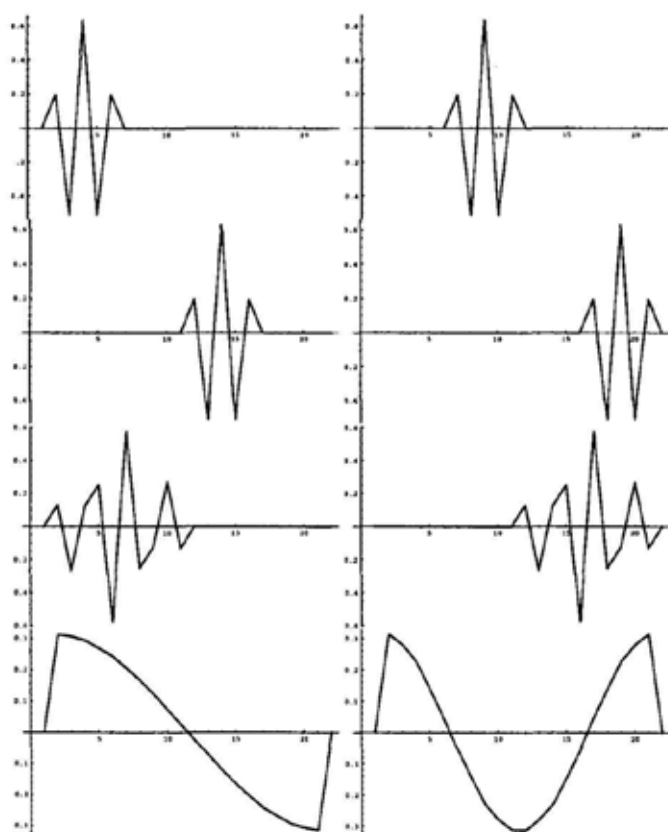


Figure 1. Construction of a basis of diadic MPCs with  $n_0 = 3$  and  $L_0 = 5$ . Left to right, top to bottom:  $\psi_{1,1}$ ,  $\psi_{1,2}$ ,  $\psi_{1,3}$ ,  $\psi_{1,4}$ ,  $\psi_{2,1}$ ,  $\psi_{2,2}$ ,  $\psi_{3,1}$ ,  $\psi_{4,1}$ . We see that equal size MPCs are identical up to a translation. Only two large scale MPCs, i.e.  $\psi_{3,1}$  and  $\psi_{4,1}$ , are shown.

a first order MPC. The other first generation MPCs, i.e.  $\psi_{1,k}$  for  $k > 1$ , are easily obtained by *translating* the support of  $\psi_{1,1}$  to the right by  $k L_0$  points (figure 1). The second step is to build the MPCs of *second generation*. They have equal width  $\ell_2$ , i.e. are twice larger, and have adjacent but disjoint supports. They are denoted by  $\psi_{2,k}$ ,  $k = 1, 2, \dots, M(2)$  and their supports are  $[(k-1)\ell_2 + 1, k\ell_2]$ . It is stressed that each MPC of second generation is *embedding* a pair of adjacent first order MPCs. We start with  $\psi_{2,1}$ , that must be normalized and orthogonal to  $\psi_{1,1}$  and  $\psi_{1,2}$ .  $\psi_{2,1}$  is therefore the solution of the optimization problem (7), using  $\phi_1 = \psi_{1,1}$ ,  $\phi_2 = \psi_{1,2}$ , and  $\phi_3 = \psi_{2,1}$ . In the terminology of section 4,  $\psi_{2,1}$  is therefore an MPC of order 3, but of second generation (the generation level is determined by the function size, whereas the order is determined by the number of constraints to be satisfied in the optimization problem). As previously, the other  $\psi_{2,k}$ s for  $k > 1$  are obtained from  $\psi_{2,1}$  by translation (figure 1).

This process is iterated in a self-similar way for MPCs of generation 3, 4, ... until the MPC of generation  $n_0$  and size  $N$  is obtained. Once the largest scale  $N$  is reached, it is necessary to build additional large scale MPCs of size  $N$  to get a complete basis of functions. Indeed,  $N$  basis vectors are needed to describe  $N$ -dimensional signals. The additional large scale MPCs are obtained by solving (7) using all vectors previously constructed as orthogonality constraints. The construction process is stopped once a total of  $N$  orthonormal MPCs is obtained.

### 5.3 Diadic MPCs with $L_0 = 2$ : degeneracy of MPCs bases

As a reference signal, we used a stationary random signal containing 20000 points, obtained by smoothing a white noise on 20 points (using a moving average). The white noise has a uniform distribution in  $[-0.5, 0.5]$ . The correlation matrix was estimated from this signal.

We first built a set of diadic MPCs with  $L_0 = 2$ ,  $n_0 = 6$  and  $n_{\max} = 7$ . We observed that these MPCs happen to be the *Haar wavelets*. Moreover, we discovered that this result is virtually independent of the reference signal considered, i.e. of the correlation matrix. In that sense, the case  $L_0 = 2$  is *degenerate*, i.e. different signals lead to virtually identical MPCs. This degeneracy is partly explained by the fact that there is little freedom in the optimization problem. Indeed, the vector  $\psi_{1,1}$ , which has two components  $(x_1, x_2)$ , must be normalized, i.e.  $x_1^2 + x_2^2 = 1$ , which leaves only one variable to adjust in the optimization. We will say that the number of degrees of freedom  $N_{df}(1)$  is one. This restricted freedom persists for MPCs of higher orders. Indeed, consider for instance  $\psi_{2,1}$ . It has four components, it must be orthogonal to two vectors (i.e.  $\psi_{1,1}$  and  $\psi_{1,2}$ ) and be normalized. It follows that  $N_{df}(2) = 4 - 2 - 1 = 1$ , i.e. there is again little room for adjustment in the optimization. More generally, it can be shown that  $N_{df}(n) = 1$  for  $n \geq 1$ .

### 5.4 Diadic MPCs with $L_0 \geq 3$ : data-adaptive MPC bases

Using the same reference signal, we built a set of diadic MPCs with  $L_0 = 3$  and  $n_0 = 6$ , so that  $N = 2^{n_0-1}L_0 = 96$ . The number of small scale MPCs (i.e.  $\ell < N$ ) is  $2^{n_0} - 2 = 62$ , and therefore the number of large scale MPCs (i.e.  $\ell = N$ ) is  $96 - 62 = 34$ , which is much larger than for the case  $L_0 = 2$ . It follows that  $n_{\max} = 62 + 34 = 96$ . We plotted only the first four large scale MPCs in figure 2. These MPCs, which are no longer similar to the Haar wavelets, were found to be sensitive to the reference signal, i.e. they are truly data-adaptive.

It is stressed that we *modified the construction rule* for large scale MPCs. At small scales, we minimized  $E((F^T \psi_{n,1})^2)$  as usual. However, at large scale, we *maximized* the mean square  $E((F^T \psi_{n,1})^2)$  instead of minimizing it. In this way, we obtain first the large scale components which carry most of the energy, as one does in a classical PC analysis. In practice, this is done simply by choosing the maximum eigenvalue solution of the system (15). We also built diadic MPCs with  $L_0 \geq 4$ . We found that they vary significantly with  $L_0$ , which indicates that a large variety of adaptive bases can be obtained with our construction method. We emphasize that small scale MPCs naturally happen to have zero or nearly zero

average. This was verified for different signals and varying  $L_0$ . In that sense, MPCs are *natural wavelets*.

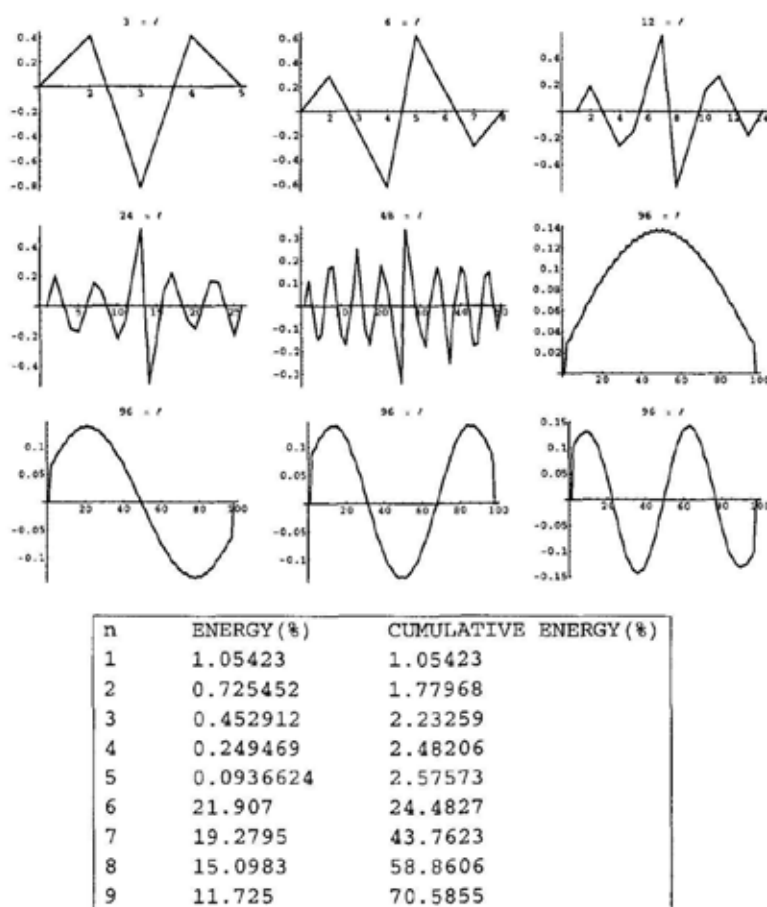


Figure 2. Basis of diadic MPCs adapted to a smoothed white noise. We used  $L_0 = 3$ ,  $n_0 = 6$  and  $n_{\max} = 96$ . Top to bottom, left to right:  $\psi_{n,1}$ ,  $n = 1, 2, \dots, 9$ . The support size  $\ell_n$  of each MPC is written on top of each plot. The large scale  $\psi_{n,1}$ s with  $n > 9$  are not shown.  $n$  is the generation order. For large scale MPCs, i.e.  $\ell = 96$ , we maximized the mean square  $E((F^T \psi_{n,1})^2)$  instead of minimizing it.

### 5.5 Example: Diadic MPCs with $L_0 = 3$ for a random binomial measure.

We have seen that the MPCs obtained with smoothed white noise are not self-similar. One may wonder if signals having a built-in self-similarity have self-similar

MPCs. To provide a partial answer to this question, we computed the MPCs for a random binomial measure.

The latter was generated by the usual dyadic self-similar cascade process. More precisely, each interval  $\mathcal{I}$  of size  $L$  is split into two adjacent disjoint subintervals of width  $L/2$ . Each subinterval receives a fraction  $w_1 = 0.3$  or  $w_2 = 0.7$  of the measure of  $\mathcal{I}$  (the measure of the initial interval is unity).  $w_1$  and  $w_2$  are chosen randomly with equal probability. This process was iterated 14 times, resulting in a signal composed of  $2^{14} = 16384$  data points (figure 3).

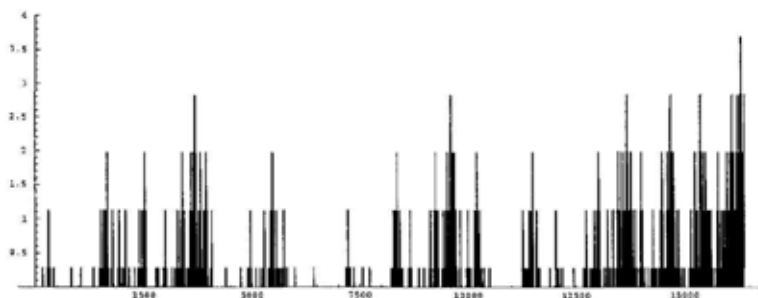


Figure 3. Logarithm of a random binomial measure.

At the end of the construction process, we took the logarithm of the resulting signal to further strengthen self-similarity. Indeed, the logarithm of a multiplicative process can be regarded as a fractal sum of pulses, where the pulses are statistically self-similar. This sum of self-similar pulses resembles a decomposition on a dyadic wavelet basis.

The MPCs obtained with this signal are shown in figure 4. As can be easily seen, the resulting MPCs are not self-similar. This means that the optimal representation of a self-similar signal, from the standpoint of MPCs, is not necessarily produced by a self-similar wavelet basis.

## 6 Conclusions

We have shown that orthonormal bases of functions with multiscale compact supports can be obtained from a generalization of principal component analysis. These functions are the eigenvectors of the correlation operator expressed in vector subspaces. Using MPCs, many approaches are *a priori* possible for the construction of orthogonal bases. In particular, we showed that we could construct dyadic wavelet bases. MPCs, which minimize their correlation with a reference signal, are data-adaptive. Moreover, MPCs are *natural wavelets*, i.e. they have typically a zero or nearly zero average. Since they minimize the energy contained in the small scale coefficients, MPCs should be very efficient for data-compression and denoising.

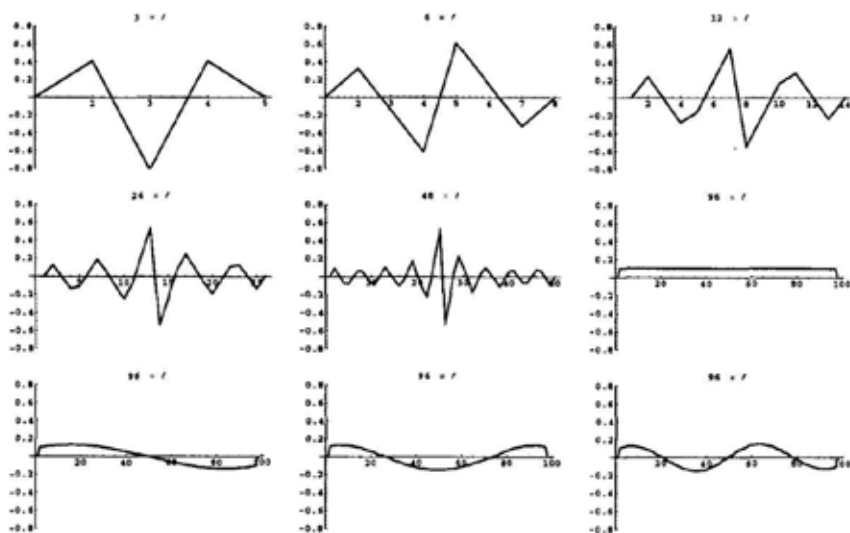


Figure 4. Basis of diadic MPCs adapted to a random binomial measure. We used  $L_0 = 3$ ,  $n_0 = 5$  and  $n_{\max} = 96$ . Top to bottom, left to right:  $\psi_{n,1}$ ,  $n = 1, 2, \dots, 9$ . The support size  $\ell_n$  of each MPC is written on top of each plot. The large scale  $\psi_{n,1}$ s with  $n > 9$  are not shown.  $n$  is the generation order. For large scale MPCs, i.e.  $\ell = 96$ , we maximized the mean square  $E((F^T \psi_{n,1})^2)$  instead of minimizing it.

## References

1. D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
2. J. S. Geronimo, D. P. Hardin, and P. R. Massopust. Fractal functions and wavelet expansions based on several scaling functions. *Journal of Approximation Theory*, 78(3):373–401, 1994.
3. R. E. Learned and A. S. Willsky. A wavelet packet approach to transient signal classification. *Appl. Comput. Harmonic Anal.*, 2(3):265–278, 1995.
4. J. M. Lilly and J. Park. Multiwavelet spectral and polarization analyses of seismic records. *Geophys. J. Int.*, (122):1001–1021, 1995.
5. S. Mallat and Z. Zhang. Singularity detection and processing with wavelets. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.
6. S. Mallat. *A wavelet tour of signal processing*. Academic Press, 525 B Street, Suite 1900, San Diego, CA 92101-4495, USA, 1998. (a) section 4.3. (b) section 7.2.1. (c) section 10.1.1. (d) section 9.1.3.
7. P. Yiou, D. Sornette, and M. Ghil. Data-adaptive wavelets and multi-scale singular spectrum analysis. *Physica D*, 142:254–290, 2000.

# COEXISTENCE OF DOUBLON AND DENDRITE STRUCTURE WITH PHASE-FIELD MODEL

Seiji Tokunaga

*Interdisciplinary Graduate School of Engineering Science, Kyushu University, Kasuga, Fukuoka, 816-8580, Japan*

*Email: tokunai@asem.kyushu-u.ac.jp*

Hidetsugu Sakaguchi

*Department of Applied Science for Electronics and Materials  
Interdisciplinary Graduate School of Engineering Sciences,  
Kyushu University, Kasuga, Fukuoka, 816-8580, Japan*

Doublon is one of the typical patterns found in crystal growth. It is a pair of symmetry broken fingers. In this paper, we obtain numerically parameter range of coexistence of doublon and dendrite structure with a phase-field model. We perform numerical simulations in a two-dimensional channel, setting small seed of crystal at left-bottom side of the channel as an initial condition. The oscillation of groove of doublon appears in some parameter range even though without perturbation. In other parameter range, both dendrite and doublon make their appearance along same growth direction.

## 1. Introduction

Crystal growth has been intensively studied as a problem of pattern formations far from equilibrium [1,2]. Many fascinating patterns such as dendrites have been studied in experiments of crystal growth [3,4,5] and computer simulations [6,7,8]. Recently, the phase-field model became one of the popular methods of computer simulations for crystal growth [9,10,11].

A diffusion field is very important in the problem of the crystal growth [2]. Doublon is one of the typical growth patterns in diffusion fields. The doublon takes a form of two fingers growing in a pair and has a narrow groove between the fingers. It has mirror symmetry with respect to the center of the groove. This doublon structure was first predicted by Ben Amar and Brener as an asymmetric dendrite along the wall of a channel [12]. They have shown analytically that the growth velocity of the doublon is in proportion to ninth power of the degree of nonequilibrium, the power of which is fairly different from the normal dendrite. The doublon patterns were found in several experiments. Akamatsu et al. found some doublon patterns in an experiment of directional solidification [13]. They confirmed that the doublon structure needs low anisotropy and high undercooling. Furthermore, they discovered that the width of groove is inversely proportional to the undercooling. Losert et al. investigated the stability of doublet structure changing the strength of fluctuation with an experiment and a phase-field model [14]. The doublon patterns were found in the experiments of Lipson's group of drying water film [15]. Doublon patterns are considered to exist in a parameter region where the dense branching morphology (DBM) appears. Ihle et al. discussed qualitatively the stability region of the doublon in the parameter space of surface tension anisotropy and the supercooling [17,18,19]. Recently, we reported the stability of doublon structure [20]. In ref. 20, doublon was classified into two, surface tension doublon and kinetic one, and each was investigated.

A phase-field model is a useful simulation model for such growth patterns that group have a form of the Ginzburg-Landau equation coupled with the diffusion equation. For melt growth, the Ginzburg-Landau equation represents the dynamics of phase transition from liquid phase to solid one and it is coupled with the heat conduction equation for the latent heat generated at the growing interface. The phase-field model is a method for numerical simulation to study interfacial pattern formation phenomena in solidification and other systems. The well-recognized appeal is to avoid the explicit tracking of macroscopically sharp phase boundaries, by introducing an order parameter  $p$ , which varies smoothly from one value to another one. It is possible to simulate the case of negligible interface kinetics that is physically relevant at low undercooling for a large class of materials, by the improvement of the model by Karma and Rappel [10].

In this paper, we use the phase-field model to investigate the detail of the transition range between doublon structure and dendrite one. We set small seed of crystal at left-bottom side of a channel as an initial condition. If one of the two fingers of doublon that steps a little ahead by some perturbation wins the competition against the other by gaining more diffusion field supply, the doublonlike structure with two symmetrical fingers will be destroyed. On the other hand, if the other finger catches up, when one finger steps a little ahead, the doublon structure is stable. We investigated the detail of the result in another paper [20]. In this paper, we will obtain a pattern coexisting of doublon and dendrite.

We introduce our model equation and the numerical method in Sec 2. In Sec. 3, we show some of the numerical results when the strength of the surface anisotropy and the degree of supercooling are changed.

## 2. Model equation

The model equations of the phase-field model are

$$\begin{aligned} \tau(\theta)\partial_t p = & \{p - \lambda u(1 - p^2)\}(1 - p^2) \\ & + \partial_x \{W(\theta)^2 \partial_x p - W(\theta)W'(\theta)\partial_y p\} \\ & + \partial_y \{W(\theta)^2 \partial_y p + W(\theta)W'(\theta)\partial_x p\}, \end{aligned} \quad (1)$$

$$\partial_t u = D\nabla^2 u + \partial_t p/2, \quad (2)$$

where,  $\lambda$  is a dimensionless parameter that controls the strength of the coupling between the phase and diffusion fields,  $p$  is an order parameter and  $p=1$  and  $p=-1$  correspond to solid and liquid phase, respectively,  $\tau(\theta)$  is an anisotropic time constant,  $W(\theta)^2$  is an anisotropic diffusion constant and  $W'(\theta)$  is a partial differential of  $W$  with respect to  $\theta$ . The variable  $u$  is the dimensionless temperature that is expressed as  $u = (T - T_M)/(L/C_p)$ , where  $T$ ,  $T_M$ ,  $L$  and  $C_p$  are respectively, the temperature, the melting temperature, the latent heat and the specific heat. The diffusion constant for  $u$  is denoted by  $D$ . The term  $\partial_t p/2$  in Eq. (2) represents latent heat production at the interface. The value  $\theta \equiv \arctan(\partial_y p / \partial_x p)$  is the angle between the direction normal to

the contours of constant  $p$  and horizontal axis.

Four-fold rotational symmetry is assumed for the anisotropy and

$$W(\theta) = 1 + e_s \cos(4\theta), \quad (3)$$

$$\tau(\theta) = W(\theta) \{1 - e_k \cos(4\theta)\}, \quad (4)$$

where the parameters  $e_s$  and  $e_k$  denote strength of surface anisotropy and that of kinetic anisotropy, respectively. Karma and Rappel derived the fundamental equation of crystal growth as the sharp-interface limit of the phase-field model as

$$\partial_t u = D \nabla^2 u, \quad (5)$$

$$u_i = -d_0(\theta) \kappa - \beta(\theta) v_n. \quad (6)$$

Equation (5) is the diffusion equation for  $u$  and Eq. (6) is the generalized Gibbs-Thomson condition, respectively, where  $d_0(\theta)$ ,  $\kappa$ ,  $\beta(\theta)$  and  $v_n$  denote respectively the anisotropic capillary length, the interface curvature, the anisotropic kinetic coefficient and the normal interface velocity. These parameters are expressed using  $W(\theta)$  and  $\tau(\theta)$  as

$$d_0(\theta) = \frac{I}{\lambda J} \{W(\theta) + W''(\theta)\}, \quad (7)$$

$$\beta(\theta) = \frac{I}{\lambda J} \frac{\tau(\theta)}{W(\theta)} \left[ 1 - \lambda \frac{W^2(\theta)}{2D \tau(\theta)} \frac{K + JF}{I} \right], \quad (8)$$

where,  $I = 2\sqrt{2}/3$ ,  $J = 16/15$ ,  $F = \sqrt{2} \ln 2$ ,  $K = 0.13604$ . If  $\tau(\theta) = W(\theta)^2$  and the parameter  $\lambda$  is chosen as  $\lambda = (2ID)/(K + JF)$ , the anisotropic kinetic coefficient  $\beta(\theta)$  vanishes, that is, the kinetic effect becomes negligible and the capillary length  $d_0(\theta) \propto 1 - 15e_s \cos 4\theta$ . If the parameter  $\lambda$  is chosen such as  $\lambda = (1.8ID)/(K + JF)$ , the anisotropic coefficients  $d_0(\theta)$  and  $\beta(\theta)$  are expressed as  $d_0(\theta) \propto 1 - 15e_s \cos 4\theta$  and  $\beta(\theta) \propto 0.1 - (e_k + 0.9e_s) \cos 4\theta$ . We use  $\lambda = (2ID)/(K + JF)$  as a typical case without the kinetic effect. We have performed numerical simulation of the phase-field model eq. (1) and eq. (2) with the finite difference method of gridsize  $\Delta x = 0.4$  and timestep  $\Delta t = 0.015$ . The simulations were done in a channel (a rectangular box) of size  $L_x \times L_y = 480 \times 96$ . We have used a channel system to study the time evolution of two fingers. (In a square box, there appear many branches and the interactions among many branches are complicated.) The initial conditions are  $p(x, y, t=0) = -1$  and  $u(x, y, t=0) = -\Delta$ , where  $\Delta$  denotes the dimensionless supercooling, except for the region of the crystal seeds. Inside of the crystal seeds,  $p(x, y, t=0) = 1$  and  $u(x, y, t=0) = 0$ . We set seed of crystal whose radius is three grids at left-bottom side of a channel as an initial condition. The boundary conditions for  $p$  and  $u$  are the no-flux boundary conditions at  $x=0$  and  $y=0, L_y$ , and the fixed boundary conditions  $p(x, y) = -1, u(x, y) = -\Delta$  at  $x=L_x$ .



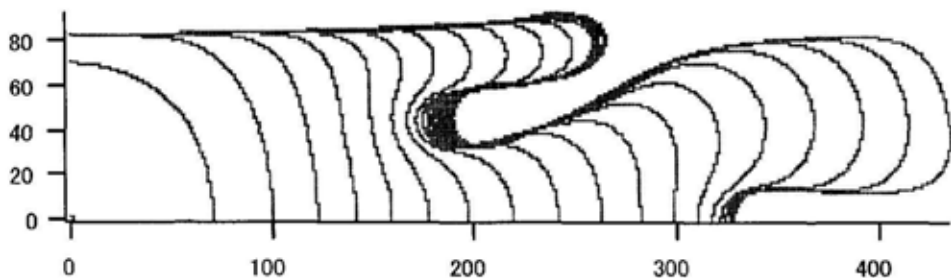
### 3. Result and discussion

We show the results of simulations in a channel with phase-field model. Figs. 1 are growing patterns increasing the strength of surface anisotropy  $e_s$  and changing supercooling  $\Delta$ . Fig. 1 (a) is a pattern in  $\Delta=0.74$ ,  $e_s=0$ . Because the anisotropy is zero, the growing crystal cannot have a clear growth direction. As a result, whole growth direction is destined to go along the channel, and a symmetry broken finger along the bottom of channel appears. When  $e_s$  is low but not zero, the main growth direction is (1,1). The crystal that is growing toward (1,1) direction forms symmetry broken pattern along the top of the wall. Fig. 1 (b) is a pattern that makes a symmetry broken finger along the top of the wall. Although Fig. 1 (c) shows the form of a symmetry broken finger like Fig. 1 (b), one finger can grow also along the bottom of wall of a channel. In fig. 1(d), a symmetry broken finger grows along the bottom of the wall, a finger runs along the top of wall. As increasing  $e_s$ , main growth direction changes from (1,1) to (1,0), therefore the symmetry broken finger tends to make its appearance along the bottom of the wall. Fig. 1(e) is a purely symmetry broken pattern growing along the bottom of channel. Fig. 1(f) is oscillating groove pattern that was reported by us [20]. In ref.20, oscillating groove pattern appears near the boundary of doublon and dendrite. Furthermore, a perturbation was adopted in the simulation of ref. 20. In this paper, we find the oscillating groove pattern between doublon and dendrite even though without perturbation. Figs.2 uses the same parameter as Fig. 1(f). If the tip of the pattern reaches  $x=440$ , the growth for the anisotropy parameter is stopped, and the numerical data for the order parameter and the temperature are saved in our computer. The order parameter and the temperature profiles in the tip region ( $p'(x,y), u'(x,y)$ ) are used for the initial conditions for the next step, that is,  $p(x,y,t=0)=p'(x+320,y)$ ,  $u(x,y,t=0)=u'(x+320,y)$  for  $x<160$ , and  $p(x,y,t=0)=-1$ ,  $u(x,y,t=0)=-\Delta$  for  $x>160$ . Fig.2(a) is a next computational process of Fig.1 (f). The oscillation is gradually attenuated and make straight groove at last in this parameter. Fig.2(b) is also a next computational process of Fig.2 (a). The groove was partially buried. In this parameter, the groove of doublon has three forms, that is, oscillating, going straight, buried. As increasing  $e_s$ , the form of dendrite appears instead of that of doublon. Fig.1 (g) is a dendritic pattern growing to (1,0) direction.

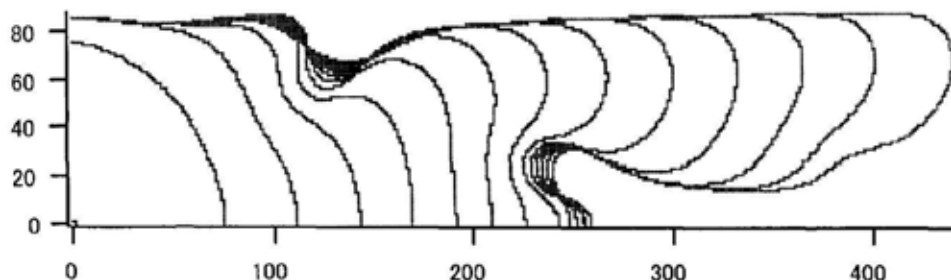
Fig.3 (a) are growing patterns fixed  $\Delta=0.8$  and  $e_s=0.007$ . Fig.3 (b) is a next computational process of Fig.3 (a) using a same method as Figs.2. In Fig.3 (a), both dendritic pattern and symmetry broken pattern can grow along the same direction. At first, the crystal which is growing along a channel is perfectly dendritic. Therefore, it seems that the dendritic pattern exists in the parameter. However, the crystal going to (1,1) direction makes a groove along the top of wall, and begins to grow. The tip of the crystal forms symmetry broken finger, that is, one of a pair of doublon. At last, since the velocity of doublon is faster than that of dendrite, doublon overtakes dendrite and dominates the diffusion field in Fig.3 (b). However, the simulation was carried out in wider channel, the dendritic pattern may preserve its appearance for longer. Though the velocity of doublon is faster than that of dendrite in this parameter range, a dendrite structure has made first. The parameter range coexisting doublon and dendrite is very narrow. Outside the parameter range, the thing made first remains to the last.

#### 4. Summary

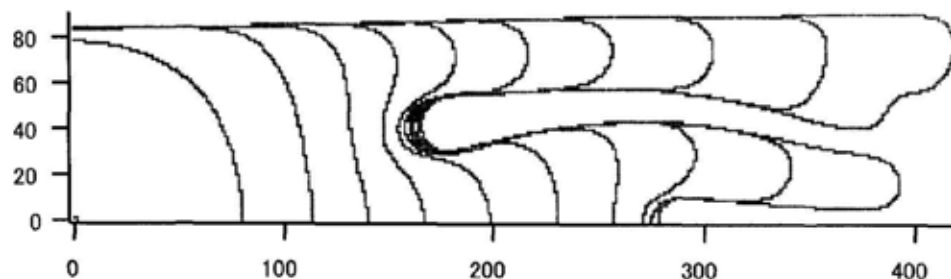
In this paper, we report some doublon patterns with phase-field model. The effect of the wall of channel is very important to make a doublon pattern. In the very narrow range, dendrite and doublon patterns coexist growing to same direction. Even though the velocity of doublon pattern is faster than that of dendrite, the dendritic pattern grows first of all.



(a)  $\Delta=0.74, es=0$



(b)  $\Delta=0.75, es=0.0035$



(c)  $\Delta=0.78, es=0.001$

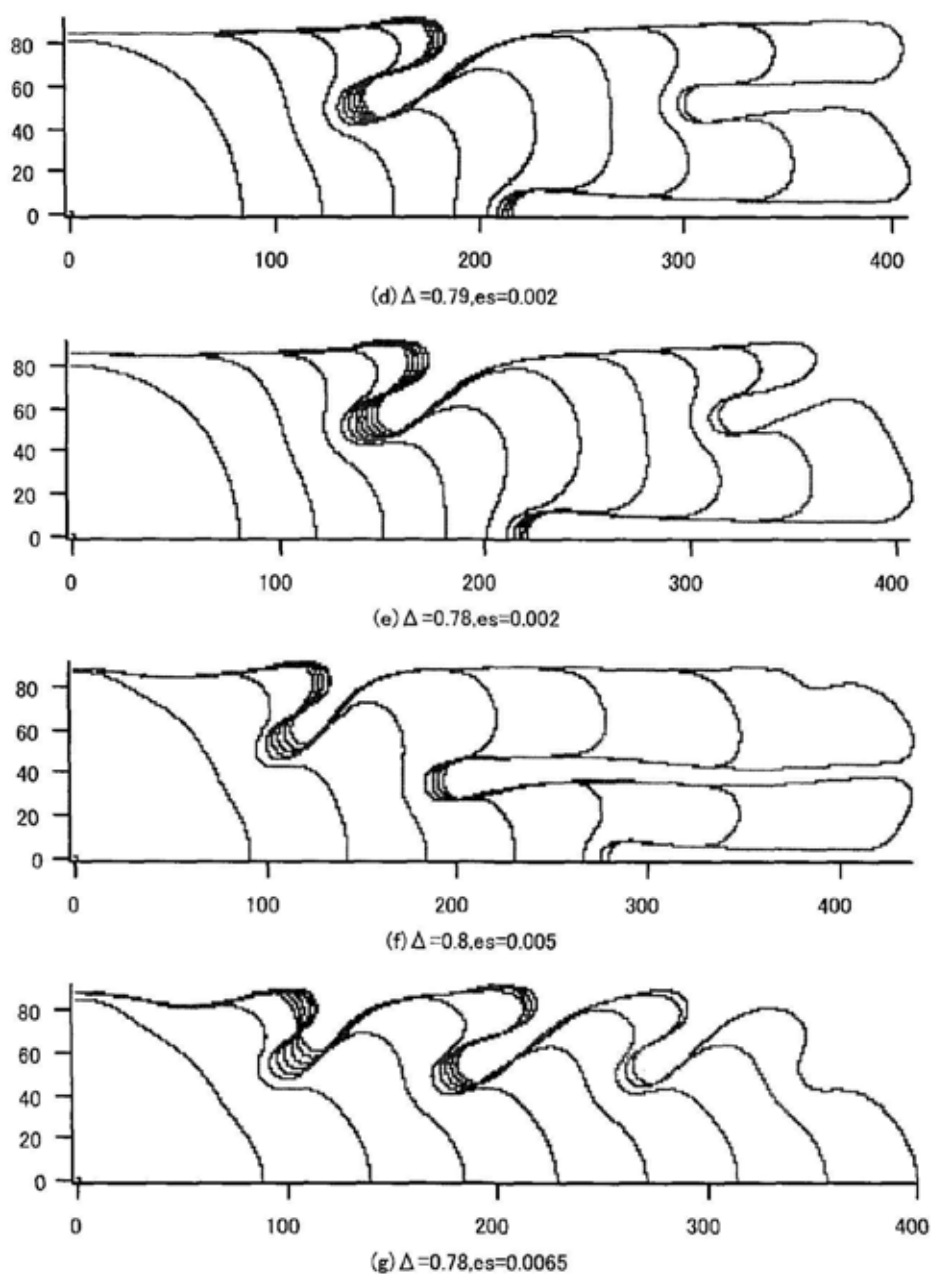


Figure 1. Growing patterns increasing the strength of surface anisotropy  $e_s$  and changing the supercooling  $\Delta$ .

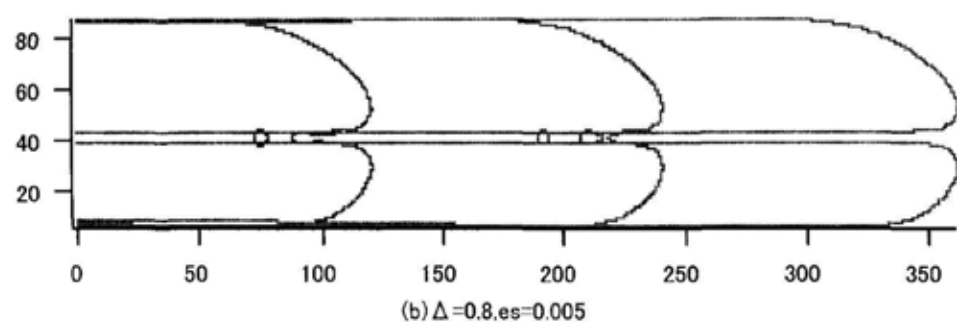
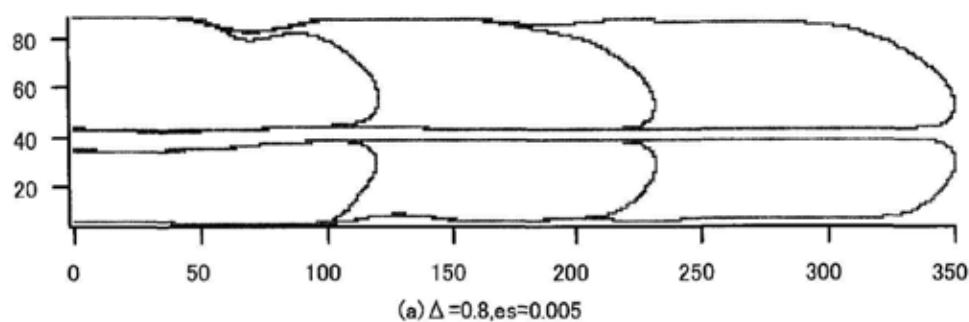
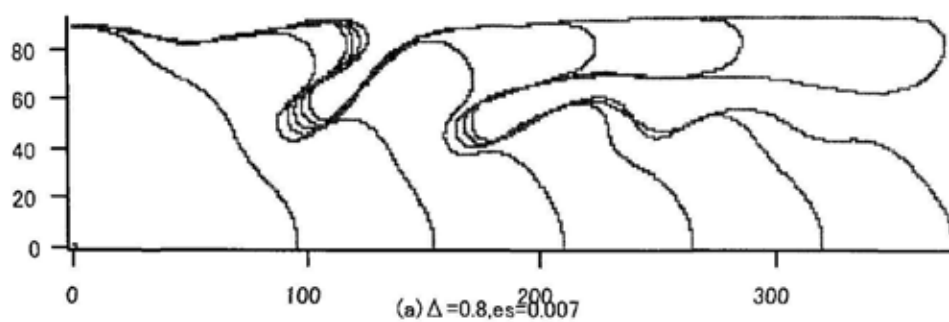
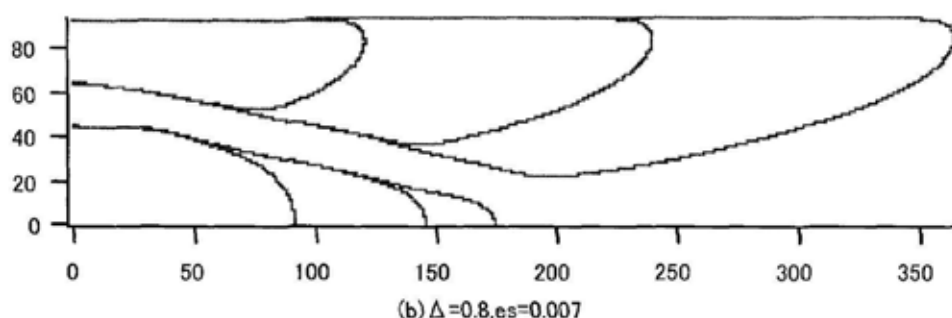


Figure 2. Second and third computational processes of Fig. 1(f).





**Figure 3.** Coexistence of symmetry broken finger and dendrite along the same direction.

## References

- [1] J.S. Langer, *Rev. Mod. Phys.* **52**(1980),1.
- [2] Yukio Saito, *Statistical Physics of Crystal Growth*. (World Scientific Publishing Co. Pte. Ltd., 1996).
- [3] S. -K. Chan, H. -H. Reimer and M. Kahlweit, *Journal of Crystal Growth* **32**, 303 (1976).
- [4] S.-C. Huang and M. E. Glicksman, *Acta Metal.* **29**, 714 (1981).
- [5] A. Dougherty, P. D. Kaplan and J. P. Gollub, *Phys. Rev. Lett.* **58**, 1652 (1987).
- [6] D. A. Kessler, J. Koplik and H. Levine, *Phys. Rev. A* **33**, 3352 (1986).
- [7] Y. Saito, G. Goldbeck-Wood and H. Müller-Krumbhaar, *Phys. Rev. Lett.* **58**, 1541 (1987).
- [8] M. Ohtaki, H. Honjo, and H. Sakaguchi, *Journal of Crystal Growth* **237**, 159(2002).
- [9] R. Kobayashi, *Physica D* **63**, 410 (1993).
- [10] Alain Karma and Wouter-Jan Rappel, *Phys. Rev. E* **53**, 3017(1996).
- [11] Alain Karma and Wouter-Jan Rappel, *Phys. Rev. E* **57**, 4323(1998).
- [12] M. B. Amar and E. Brener, *Phys. Rev. Lett.* **75**, 561(1995).
- [13] Silvere Akamatsu, Gabriel Faivre, and Thomas Ihle, *Phys. Rev. E* **51**, 4751(1995).
- [14] W. Losert, D. A. Stillman, H. Z. Cummins, P. Kopczynski, W.-J. Rappel and A. Karma, *Phys. Rev. E* **58**, 7492 (1998).
- [15] N. Samid-Merzel, S. G. Lipson and D. S. Tannhauser, *Physica A* **257**, 413 (1998).
- [16] H. Sakaguchi and S. Tokunaga, *Prog. Theor. Phys.* **109**, 43 (2003).
- [17] T. Ihle and H. Müller-Krumbhaar, *Phys. Rev. E* **49**, 2972(1994).
- [18] E. A. Brener, *JETP*, **82**, 559 (1996).
- [19] E. Brener, H. Müller-Krumbhaar and D. Temkin, *Phys. Rev. E* **54**, 2714(1996).
- [20] S. Tokunaga and H. Sakaguchi, submitted to *Phys. Rev. E*.

## FRACTALITY AND FRACTAL DIMENSION IN MESOAMERICAN PYRAMID ANALYSIS

GERARDO BURKLE-ELIZONDO

*Universidad Autónoma de Zacatecas. Centro Interinstitucional de Investigaciones en Arte y Humanidades. Unidad de Postgrado II. CP 98060. Zacatecas Zac. México*  
E-mail: burklecaos@hotmail.com

ANA GABRIELA FUENTES-LARIOS

*Universidad del Valle de Atemajac. Facultad de Administración y Turismo. Tepeyac # 4800. CP 45050. Zapopan. Jalisco. México*  
E-mail: gfuente@cybercable.net.mx

RICARDO DAVID VALDEZ-CEPEDA

*Universidad Autónoma Chapingo. Centro Regional Universitario Centro Norte. Apdo. Postal 196, CP 98001, Zacatecas Zac. México*  
E-mail: vacrida@hotmail.com

The proofs seems to show that the ancient Mesoamerican architects and artists developed geometrical concepts, and used them in their works, for example to get orientations of the buildings with a relationship with the geomancy and the alignment with the equinox, as was signed by Aveni, Hartung and Broda.<sup>1,2,3</sup>

Pyramids had a religious function related to the myth and the ritual expressions, traditions and ideology. These buildings were established in specific sacred spaces in order that the people can experience a powerful holy event. The relationship between death, art and architecture is also evident<sup>4</sup>. A lot of ritual ceremonies take place on the top of the mountains or pyramids or on the platforms like steps. A pyramid was fundamentally a ceremonial building that represented a pattern of the cosmological organization and the center of the world<sup>5</sup>.

A pyramid is a series composed by different number of platforms of different sizes. Then we have to analyze all the structure together, but by the other side, we have to see these buildings like boundaries and try to study their separated sequential segments in order understand better the distinct aspects of the correlation functions.

The aim of this work is to study these structures trying to find out the patterns and designs and the forms into this complex geometry that appear to enclose a specific guide of information encode in them. What we want to decipher the possible interconnected nature of different reckoning systems. To do it we present here three different procedures of analysis. The first one studies the structures like series from the point of view of areas against volumes. In a second one we visualize the pyramid like the reason of the volume interpolated with its empty complement mould. The third one is the calculation of the fractal dimension of a big number of pyramids with the Box counting method that shows rather the roughness of an object or fluctuations of the height over length scale. In the past we found those Mesoamerican artworks, sculptures and architecture to have fractal dimension.<sup>6</sup>

In the first group we included 16 pyramids to study. The Fractal Dimension average was  $1.236 \pm 0.108$  with  $r^2 = 0.918$ .

We name the second procedure "reason of the volume against the empty complement". To get logic structure it is necessary to add an extra imaginary platform at

the top in order to form a mould, what makes it a complex model. The analysis was done with the GS+ program (version 5.1.1 Gamma Design software, 2001). This model can be described through the following equation:

$$\sum_{i=1}^n VP_i \propto (VT - \sum_{i=1}^n VP_i)^D$$

Third procedure. Fractal Dimension. In these group 14 pyramids were included in the analysis. The Fractal Dimension general average of this group was  $1.312 \pm 0.179$  with  $r^2 = 0.874$ . We collected 26 images from pyramids of different Mesoamerican cultures that were scanned and saved as bitmap files on a computer. Thereafter the images were analyzed with the program Benoit, version 1.3<sup>7</sup> in order to calculate Box, Information and Mass Dimension, and their intercepts on log-log plots.

The total averages of this group were for Box Dimension  $Db = 1.931 \pm 0.010$ , Information Dimension  $Di = 1.941 \pm 0.00017$  and Mass Dimension  $Dm = 1.959 \pm 0.042$ .

The first two procedures show the existence of fractality in series of different kind of pyramid measurements and the change or relation between the parameters.

The reason as to why in the third procedure we got bigger Fractal Dimensions than in the two first, could be due to this one measuring the roughness of an object as a whole, what is completely different to at measuring geometric relations of particular scalar properties and study the power function describing the fractality of the pyramids.

## References

1. Aveni Anthony F. Observadores del Cielo en el México Antiguo. Fondo de Cultura Económica. México, 1997.
2. Hartung Horst. Arquitectura y Planificación entre los Antiguos Mayas: Posibilidades y Limitaciones para los Estudios Astronómicos. Astronomía en la América Antigua. En: Aveni A.F. Siglo XXI Editores. 1980, págs. 145-167.
3. Broda Johanna. Arqueoastronomía y Etnoastronomía en Mesoamérica. Universidad Nacional Autónoma de México. México 1991.
4. Matos Moctezuma Eduardo. The Great Temple of Tenochtitlan: Model of Aztec Cosmovisión. In Mesoamerican Sites and World Views, Elizabeth P. Benson ed. Dumbarton Oaks, Washington D.C. 1981, pages 71-86.
5. David Freidel. Schele Linda. Parker Joy. El Cosmos Maya. Fondo de Cultura Económica. México, 1999.
6. Burkle Elizondo Gerardo. Valdez Cepeda Ricardo David. Do The Artistic and Architectural Works Have Fractal Dimension? In: Emergent Nature. Patterns, Growth and Scaling in the Sciences. Miroslav M. Novak Ed. World Scientific. 2001, pages 431-432.
7. TruSoft Int'l Inc. Benoit, version 1.3: Fractal Analysis System. (20437<sup>th</sup> Ave. No. 133, St. Petersburg, Fl 33704, USA).
8. Marquina Ignacio. Arquitectura Prehispánica. Instituto Nacional de Antropología e Historia. Secretaría de Educación Pública. México, 1990.

## MORPHOLOGICAL VARIETY IN CRYSTAL GROWTH OF MERCURY (II) CHLORIDE ON AGAR SLIDES

J. ALBERTO BETANCOURT-MAR AND E. JONATHAN SUÁREZ-DOMÍNGUEZ

*Universidad del Noreste, Prol. Ave. Hidalgo 6315, Col. Nuevo Aeropuerto, Tampico, Tams.,  
México, C.P. 89337. Email: nolineal@spectroq.com*

It is well known that crystals of some substances can grow in dendritic patterns. The morphology exhibited by crystals of certain substances that grow in the surface of a thin agar plate is complex and can be described by fractal geometry [1,2,3,4]. Some of these structures can be explained by the Diffusion Limited Aggregation model (DLA) [5]. However, there are other structures as dense branching morphology that need different models [2].

In this work, the morphology variety of crystal growth of mercury (II) chloride (Merck) in thin agar-agar (Merck) plates was studied. To carry out the experiments it was necessary to divide all the microscope slides used into 4 equal parts of 1.0 X 1.5 cm each, edging the divisions of each slide with silicon glue. These four spaces in the slides have been called cells. Experiments were performed by spreading 25  $\mu$ L or 50  $\mu$ L of the solutions on the cells. It was observed that when water evaporates crystals start to develop. Drying times and temperatures (30 to 70 °C) were controlled. The center of cells took one minute to 5 hours to dry, this time is called drying time.

Finally, the crystals obtained after the evaporation of solution, were observed through a microscope at 50 and 100 magnifications. They were also photographed using a Samsung Digimax 101 camera at its highest resolution.

At temperatures of  $30 \pm 2$  °C, it was observed that the morphology depends more strongly on the drying time than on the concentration. At long drying times, the structures are DLA-like. The fractal dimension was determined by box counting method [2] and by mass-radius method [2] and the results agreed with the values reported in literature (around 1.7) [2,5].

At short drying times compact structures and patterns of crystallization appear in bands. These bands can be considered as periodic crystallization (Fig. 1a).

The same tendency occurs at temperatures of  $50 \pm 5$  °C and  $70 \pm 5$  °C, though new patterns were observed, as can be seen in Fig. 1b, that shows spirals of crystallization that grow inside, towards a central compact crystal.

At temperatures of  $30 \pm 2$  °C, the spirals do not form, but sometimes a ring of crystallization around a big compact crystal can be observed.

The DLA-like structures are a well known phenomenon in crystallization [2,5]. DLA-like structures appears at long drying time (not too long, because of the appearance of faceted crystals instead) because this process is governed by diffusion. At short drying times, the velocity of growth is higher and therefore, the crystallization is more irregular: the degree of randomness is higher. These structures are thinner than DLA-like (their color is paler). That is: the structures are more random, flatter, they do not have enough time to crystallize in ordered and thicker structures.

Although the bands resemble Liesegang rings, they are not the same because in the actual experiments the initial concentrations of ions are homogeneous and the bands of crystallization are approximately equidistant. This could be explained as follows: initially, the concentration of the salt is homogeneous, however, when the water evaporates, the concentration increases following an evaporation front (the edge of the crystallization and the moist gel). If the evaporation is fast the salt will supersaturate. Suddenly, some of the salt crystallizes and these crystals consume the ions around it. Here a band of crystal appear followed by an empty space. The evaporation front goes ahead and the ions can



not diffuse to empty space and in this space the concentration never reaches high enough level for the salt to crystallize. Nevertheless after the empty space there are ions at higher concentration that can be even more concentrated until supersaturation. The process is repeated until the velocity decreases or the evaporation front stops.

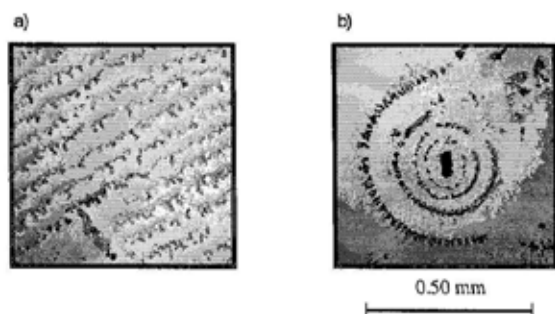


Fig. 1 Example of periodic and spiral crystallization. a) was obtained by growing 25  $\mu\text{L}$   $\text{HgCl}_2$  0.20 M in agar 0.75 g/100 mL at  $25 \pm 2^\circ\text{C}$ , drying time was 5 min; b) was obtained by growing 50  $\mu\text{L}$   $\text{HgCl}_2$  0.10 M in agar 0.20 g/100 mL at  $50 \pm 5^\circ\text{C}$ , drying time was 10 min. The scale is the same for both images.

At higher temperatures spirals can appear. The explanation could be the same: sometimes, a crystal grows slowly before it is touched by an evaporation front. When one or more evaporation fronts move towards this crystal, the crystallization stops in some region around the previous crystal (the salt around it has been depleted). Then, empty regions with a crystal in the center can be seen. The fronts can generate concentric circles around the central crystal (by the same mechanism as the bands of crystallization). At higher temperatures, the growth velocity is higher too, and possibly there are different velocities for the evaporations fronts and these velocities are changing over time; this way, instead of concentric circles, a spiral appears.

More research is needed to prove and refine the mechanisms proposed above for bands and spirals in crystal growth on agar slides. Maybe this mechanism can be applied to other phenomena with periodic crystallization.

## References

1. Mandelbrot, B. B. *The Fractal Geometry of Nature*. Freeman. 1983
2. Vicsek, T. *Fractal Growth Phenomena*. Second Edition. World Scientific, 1992
3. Yasui, M.; Matsushita, M. "Morphological Changes in Dendritic Crystal Growth of Ammonium Chloride on Agar Plates". *Journal of The Physical Society of Japan*. **61** [7] 2327-2332. 1992
4. Suda, J.; Matsushita, M.; Izumi, K. "Morphological Diversity in the Crystal Growth of Potassium and Rubidium Dichromates in Gelatin Gel". *Journal of The Physical Society of Japan*. **69** [1] 124-129. 2000
5. Witten, T.A; Sander, L. M. Diffusion limited aggregation, a kinetic critical phenomenon. *Phys. Rev. Lett.* **47** [19], 1400-1403. 1981

# **FRACTAL CHARACTERISTICS OF BAINBRIDGE CRATER LAKE SEDIMENT GRAY-SCALE INTENSITY DATA DOCUMENTING THE FREQUENCY AND INTENSITY OF HOLOCENE EL NIÑO/SOUTHERN OSCILLATION EVENTS**

NATALIA A. BRYKSINA

*Institute of Mineralogy and Petrography, Pr. Ak. Koptuyuga 3, Novosibirsk 630090, Russia  
E-mail: bryksinan@cc.umanitoba.ca*

WILLIAM M. LAST

*Department of Geological Sciences, University of Manitoba, 125 Dysart Road, Winnipeg,  
Manitoba R3T 2N2, Canada  
E-mail: WM\_Last@UManitoba.ca*

El Niño/Southern Oscillation phenomenon is regarded as one of the most important elements in year-to-year variations in the Earth's climate, with over half of the globe being subjected to weather anomalies associated with this phenomenon [2]. The Galápagos Islands, located within the core of the El Niño/Southern Oscillation region, contain numerous closed-based saline lakes that are ideally situated to provide a potentially continuous, long-term record of ENSO events. Bainbridge Crater Lake, occupying a small eruption cone, is a ~3 m deep hypersaline lake. The lithostratigraphy and mineralogy of a 4.1 m long sediment core from the offshore portion of the basin was used to investigate the hydrological and brine chemical fluctuations over the past 6200 years by Riedinger et al. [5]. X-radiography of the core was used in this study to better define the frequency and intensity of El Niño events. Last et al. [4] provide methodological details on x-ray image acquisition, image enhancement and analyses. More than 500 distinct lamination couplets (e.g., light-dark beds), representing individual ENSO events, identified on the x-radiographs provide the basis for our assessment of millennial-scale variability of El Niño over the past six millennia.

Fractal geometry has been applied to a wide range of phenomena in recent years [3]. Many observations of nature consist of records in time or a series of observations, which can be characterized by the Hurst exponent ( $H$ ). The trace of the record is a curve with fractal dimension  $D=2-H$ ,  $0<H<1$ . When the Hurst exponent  $H$  is greater than 0.5, the series record is persistent: that is, an increasing or decreasing trend in the past favors an increasing or decreasing trend in the future. The increments are positively correlated. If  $H < 0.5$ , the increments are negatively correlated. For ordinary Brownian motion,  $H=0.5$ .

Here we analyze in terms of Hurst exponent and fractal dimension the variability in gray level of x-radiographs of finely-laminated sediments from Lake Bainbridge Crater, Galápagos. We obtain fractal characteristics of gray-scale intensity data for two parts of sedimentation (Part1: relatively recent 3 millennia, Part2: the period from 3-6 kyr B.P.) and compare them. Considering the sedimentation rate is approximately constant on these two parts of the stratigraphic sequence, gray-scale intensity data series are approximately analogous to a time series with 0.2  $^{14}\text{C}$  year-interval for Part1 and with 0.4  $^{14}\text{C}$  year - interval for Part2. The dark-light stratigraphic variation evident in the x-radiographs represents fluctuation from strong El Niño (low gray-scale intensity values) to non-El Niño (high gray-scale intensity values).

Two methods have been used to measure the Hurst exponent of gray-scale intensity data: the power spectrum method and the width method. The power spectrum method gives the fractal characteristics of data as a whole, and the width method characterizes them depending on length or time scale.

The Hurst exponents found by the power spectrum method are equal to 0.88 and 0.8 for Part1 and Part2 data respectively. These rather high values of Hurst exponent indicate that gray-scale intensity data as whole have persistent behavior in both parts of stratigraphic sequence. Persistence means that these data exhibit clear tendency with relatively little noise. Non-randomness of grayscale intensity data is characterized by a fractal dimension ( $D=2-H$ ), which is equal to 1.12 for data from Part1 and 1.2 for data from Part2. That means that gray-scale intensity data are strongly non-random. We interpret this persistent stratigraphic behavior in the distribution of laminae as reflecting hydrologic control by the external influence of El Niño/Southern Oscillation phenomena.

The values of the Hurst exponents obtained by the width method vary from 0.09 to 0.77 depending on time scale and, therefore, gray-scale intensity data have different behaviors on different scale. The youngest part of the record has a persistent behavior with  $H=0.77$  for times up to about 5.6 years and the older part of the sequence has persistent behavior with  $H=0.64$  for times up to about 26 years. Decreasing time scale of persistent behavior of gray-scale intensity data in more recent sedimentation record indicates that data from Part1 exhibit clear tendency, for example periodicity, on smaller scales than data from Part2. As any periodical function has persistent behavior on length scale equaled to half of its period, then we may suggest that data from Part1 have periodicity about 11.2 years and data from Part2 about 52 years. This result is consistent with other conclusions in literature about increasing frequencies of ENSO activity since 3000 yr B.P[5].

A persistent behavior with the Hurst exponent of  $H=0.7$  was also found in the analysis of gray-scale intensity data of oscillatory zoning in cave calcite from Hungary (cave Semlo-Hed) [1]. Formation of oscillatory zoning in this calcite was also controlled by external processes occurring outside of the cave.

## References

1. Bryxina N. A., Dublyansky Yu. V., Halden N. M., Campbell J. L. and Teesdale W. J., Statistical characteristics of oscillatory zoning in cave calcite - popcorn from Hungary: *Dokl. Akad. Nauk.* **372** (2000) pp. 514-517.
2. Diaz H. F. and Maekgraf V. El Niño Historical and Paleoclimatic Aspects of the Southern Oscillation (Cambridge University Press, Cambridge, 1992).
3. Feder, J. Fractals. (Plenum Press, New York, 1988).
4. Last W. M., Bryxina N. A., Baxter K., Riedinger M. and Mark Brenner, Mineralogy, geochemistry and event stratigraphy of Bainbridge Crater Lake, Galápagos Islands. (International Limnogeology Congress, Tucson, Abstracts Volume, 2003) pp. 153-153.
5. Riedinger M. A., Steinitz-Kannan M., Last W. M. and Brenner M., A ~6100  $^{14}\text{C}$  yr record of El Niño activity from the Galápagos Islands: *Journal of Paleolimnology*, **27** (2002) pp. 1-7.

## FRACTALS AND PLANT WATER USE EFFICIENCY

A. BARI & G. AYAD

*International Plant Genetic Resources Institute, Via dei Tre Denari, 00057 Maccarese, Italy*  
E-mail: [a.bari@cgiar.org](mailto:a.bari@cgiar.org) & [g.ayad@cgiar.org](mailto:g.ayad@cgiar.org)

A. MARTIN & J.L. GONZALEZ-ANDUJAR

*Institute of Sustainable Agriculture, Cordoba, Spain*  
E-mail: [gelmamua@uco.es](mailto:gelmamua@uco.es) & [andujar@cica.es](mailto:andujar@cica.es)

M. NACHIT & I. ELOUAFI

*CIMMYT, c/o International Center for Agricultural Research in the Dry Areas, Syria*  
E-mail: [m.nachit@cgiar.org](mailto:m.nachit@cgiar.org) & [i.elouafi@cgiar.org](mailto:i.elouafi@cgiar.org)

Plant species adapt themselves to a wide range of water availability conditions from wet to desert areas. Most plants lower their  $K_e$  (water requirements) [1] either through conservation mechanisms or by enhancing their uptake of limited soil moisture.  $K_e$  is the ratio of the actual optimal water absorbed by a plant ( $ET_{crop}$ ) to a reference  $ET_0$  called potential evapo-transpiration, which is calculated from meteorological station data.

$$ET_{crop} = K_e \times ET_0 \quad (1)$$

The depth and density of the root system are considered to be major factors determining the amount of water absorbed by a plant. Deep taproots in combination with shallow surface roots permit plants to capture moisture effectively, such as that from light rains or from lower parts of the soil profile. Quantification of such root architectural traits is difficult, and describing root systems based on their biomass or length distribution has not proven to be completely successful.

The use of fractals, however, to study the root branching of plant species such as sorghum, has revealed that the genotypes of African origin are more highly branched, with deep roots, than US-derived genotypes [2]. The fractal dimensions of plant roots have also been found to differ among genotypes of the common bean [3]. Earlier work by Tatsumi and Takagai [4] found fractals useful for diagnosing root development, and revealed that the fractal dimension can be a good indicator for estimating system size as well as the complexity of root branching.

The aim of the current study is to assess genotypic variability ( $K_g$ ) of root branching in olives (*Olea europea* L.), as per the new proposed Equation 2 that takes into account the variability within the olive crop and water requirements. More specifically, we seek to determine quantitative architectural/branching parameters using fractals along with analysis of physiological processes underlying genotypic WUE such as water flow and transpiration of the plant.

$$ET_{crop} = K_g * K_e \times ET_0 \quad (2)$$

Pencil-long, pencil-thick cuttings were taken from five different olive cultivars (G1–G5). Prior to planting, cuttings were dipped in rooting hormone to improve the strike rate, and the rooting was then carried out under mist. At 3 months old, the plants were placed on a copy stand and their images were captured using a digital camera linked to a computer. The root fractal dimension  $D$  parameters ( $D_b$  and  $D_r$ ) were determined for 30 root images. The fractal dimension values of the roots were then contrasted against another set of field data related to the measurement of stomatal conductance (porosity), stomatal resistance,

and transpiration from the leaves. The stomatal properties are related to the ease with which water vapor diffuses out through small pores (stomata) in the leaves during transpiration and the ease with which carbon dioxide diffuses in through the same pores during photosynthesis (carbon fixation). The measurements were taken on olive trees using an LI-1600 portable infrared gas analyzer (LI-COR, Inc., USA).

The highest values of the different fractal dimensions of the roots were found in cultivar G3. In terms of results related to the WUE parameters measured on the leaf, the same cultivar (G3) had the lowest stomatal conductance, with a value of 0.36. In terms of water flow into the plant, G3 had the highest value of 4.5. Intriguingly, the same cultivar, G3, loses less water, having a transpiration value of 6.37, the lowest among all five cultivars (Table 1).

**Table 1.** WUE parameters and root fractal dimension values for each olive cultivar

Cultivar	Water flow (kg s <sup>-1</sup> )	Transpiration (kg s <sup>-1</sup> m <sup>-2</sup> )	Stomatal conductance (mmol m <sup>-2</sup> s <sup>-1</sup> )	$D_b$ (edge)	$D_r$
G1	3.33	8.20	0.50	1.889	1.064
G2	2.83	8.47	0.54	1.910	1.146
G3	4.50	6.37	0.36	1.919	1.194
G4	2.23	6.67	0.41	1.903	1.189
G5	3.83	9.03	0.58	1.907	1.181

The analyses conducted demonstrate that fractals can be used effectively to discriminate between cultivars and, most importantly, may also assist in the selection of cultivars with superior WUE.

We are currently investigating the measurement of root branching/architecture through fractals in further detail in relation to WUE parameters, and will carry out genomics studies to shed light on the inheritance—and, in particular, the function—of root architecture in relation to water use efficiency.

## References

- [1] FAO: Crop evapotranspiration - Guidelines for computing crop water requirements FAO Irrigation and drainage Rome, 1998.
- [2] Masi C.E.A. and Maranville J.W., Evaluation of sorghum root branching using fractals, *Journal of Agricultural Science* **131**(3) (1998) pp. 259–265.
- [3] Nielsen K.L., Miller C.R., Beck D. and Lynch J.P., Fractal geometry of root systems: Field observations of contrasting genotypes of common bean (*Phaseolous vulgaris* L.) grown under different phosphorus regimes, *Plant and Soil* **206** (1999) pp. 181–190.
- [4] Tatsumi J. and Takagai K., Fractal characterization of root system architecture in legume seedlings. In M. M. Novak and T. G. Dewey (eds.) *Fractal Frontiers*, World Scientific, Singapore, 1997.

# NEED AND FEASIBILITY OF APPLYING L-SYSTEM MODELS IN AGRICULTURAL CROP MODELING

L. PACHEPSKY, M.KAUL, CH.WALTHALL AND C. DAUGHTRY

*Hydrology and Remote Sensing Laboratory, ARS-USDA, 10300 Baltimore Ave, Beltsville, MD, USA  
E-mail: lpachepsky@hydrolab.arsusda.gov*

J. LYDON

*Sustainable Agricultural Systems Laboratory, ARS-USDA, 10300 Baltimore Ave, Beltsville, MD,  
USA E-mail: lydonj@ba.ars.usda.gov*

Development of open parametric L-systems creates an exiting prospect of crop modeling visualization; this allowed the explicit effects of environment on the L-model.

Two varieties of soybean, Essex (a conventional grain type) and Moon Cake (a tall growing vegetable type) were growing in three controlled climate chambers at a photoperiod 14 hours, light intensity  $390 \mu\text{mol m}^{-2} \text{s}^{-1}$ , and temperatures 32/27, 26/21, and 20/15°C (day/night). Temperature and photoperiod are the leading environmental variables determining the rate of progress towards flowering for soybeans [1]. The models for quantitative description of soybean vegetative development were taken from [2]. For visual modeling, a software L-Studio [3] was used.

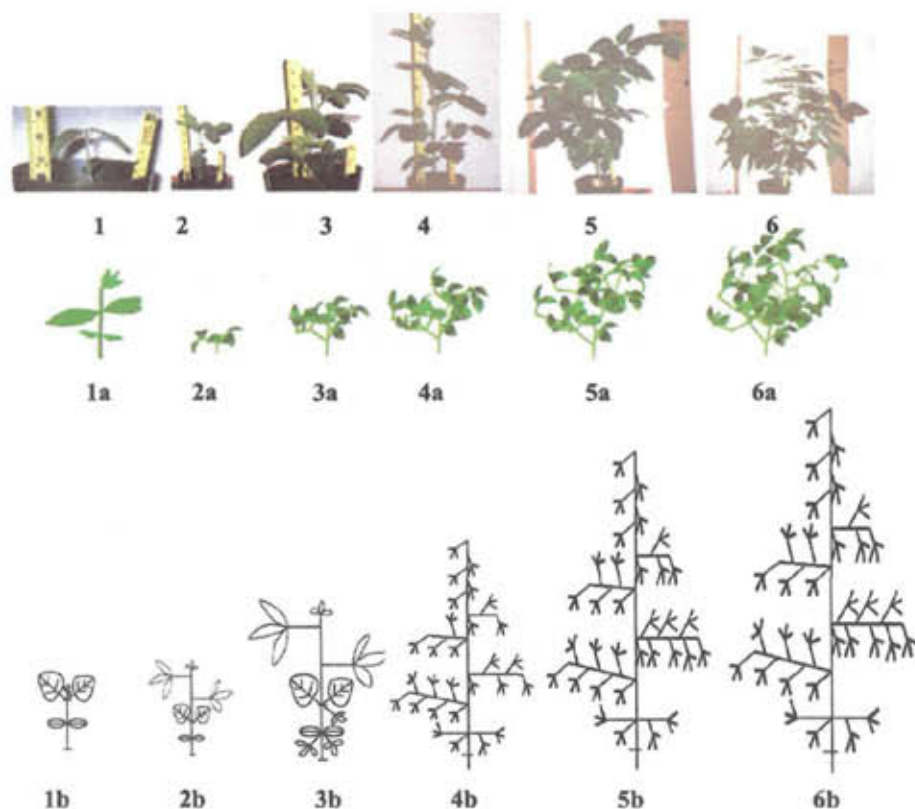


Figure shows photographs (1-6) of the plants of cultivar Essex at the moments 6 series of measurements for temperature 32/27°C, a visual L-model of these plants (1a-6a), and plant maps (schematically) at the same moments (1b-6b). The maps were used for recording the measurements and as a first step of data generalization for L-modeling.

There was no qualitative significant morphological difference between two cultivars. The effect of temperature was significant from the moment of emergence. Simulations were successfully run for all treatments.

Specifics of applying L-systems in crop modeling consists in the fact that most of crop models simulate a "typical" or an "average" plant in a canopy. For such plant, a mechanistic crop model (i.e. GOSSYM, GLYCIM) provides information on internode elongation rates, rates of leaf appearance, growth of leaf area, branching, leaf turgor and senescence and biomass distribution between organs as dependent of environmental variables. Data collection, that is sufficient to parameterize such model, is sufficient also to parameterize the L-system model. Linking a mechanistic crop simulator with L-system appears to be feasible.

In general, the studies at the level of individual plant are not given the attention they deserve, and the wider use of L-system modeling can help mend this situation. However, even at the level of a single plant the issue of accuracy depends on the question asked. For example, it is not obvious that estimating light interception by leaves of different age may require the same accuracy of the plant architecture representation as estimating temperature and gas regime and gradients. In crop modeling, where one deals with a canopy with high variability in individual plant parameters, architecture representation with L-systems may require yet different accuracy. Here the evaluation of L-system model accuracy should be what Wösten et al. [4] called functional.

Introduction of L-systems in crop modeling would add a new facet to the problem of crop model validation. The analysis of the basics of the user interface requirements [5] shows that the L-systems model coupled with a crop model could serve as an interactive and attractive for users component of a crop model interface.

1. Summerfield R.J. et al., Towards the reliable prediction of time to flowering in six annual crops: Soybean (*Glycine max*). *Expl. Agr.* 29 (1993) pp. 253-289.
2. Acock, B., Pachepsky, Y.A., Acock, M.C., Reddy, V.R., Whisler, F.D., Modeling soybean cultivars development rates, using field data from the in Mississippi Valley. *Agron. Jour.* 89 (1997) pp. 994-1002.
3. Měch, R. et al. *CPFG*. Version 3.4. User's Manual (1998).
4. Wösten, J. H. M., C. H. J. E. Schuren, J. Bouma, and A. Stein.. Use of practical aspects of soil behavior to evaluate different methods to generate soil hydraulic functions. *Hydrol. Processes* 4 (1990) pp. 299-310.
5. Acock, B., Pachepsky, Y.A., Mironenko, E. V., Whisler, F.D. Reddy V.. GUICS: A generic user interface for on-farm crop simulation. *Agron. Jour.* 91 (1999) pp. 657-665.



# **FRACTAL DETECTION AND AVOIDANCE USING RS STATISTICS AND HONEYBEE NAVIGATIONAL SKILLS IN DYNAMIC ENVIRONMENTS**

Reginald L. Walker  
Tapicu, Inc., P.O. Box 88492  
Los Angeles, CA 90009, USA  
E-mail:rwalker@tapicu.com

The nondeterministic time-invariant statistics<sup>1</sup> needed to understand and model the nature of the distribution of Internet transmissions associated with file transmissions result in heavy-tailed distributions which reflect their fractal-like behavior<sup>2</sup>. The fractal-like behavior implied that the burstiness (degree of self-similarity) of aggregated Ethernet traffic (LAN or local area networks) should not be modeled by pure Poisson or Poisson-related models. The Poisson type models were shown to be less bursty as the number of traffic sources increased. The fractal nature of data transmissions for the Internet results from a diverse and random set of transmission patterns for file transfers. Some of the self-similarity in Internet traffic can be attributed to computer file system characteristics and user behavior.

This paper presents the Tocator Apicu information sharing (IS) model<sup>3</sup> which benefits from extending the view of the World Wide Web ( $W^3$ ) as an information universe<sup>4</sup> that contain thematically unified clusters (TUCs)—random collection of websites<sup>2</sup>, and incorporates the honeybee information sharing model. The information sharing model uses an information ecosystem that maps the Internet to a composite set of self-contained ecosystems (TUCs and ISPs—Internet service providers). Information availability of each self-contained ecosystem reflects stochastic fluctuations<sup>2</sup> that can occur within randomly selected areas of the Internet which can be detected by measuring the self-similarity (fractal) behavior in the Web. The network factors that result in the fractal behavior of Internet traffic patterns were detected by using *rescaled adjusted range* or *RS statistics* to individually measure the degree of self-similarity (fractal behavior) between the location of the Tocator Apicu HTML Resource Discovery (HRD) system and selected TUCs using the global Internet navigational backbone.

The information ecosystem merges the social hierarchy of honeybees information sharing with the techniques of distributed and scientific computing, high-performance knowledge discovery in databases (KDD), hypertext, information retrieval (IR), and wide-area networking—the Internet. As in each localized view of an honeybee colony's ecosystem, each self-contained information ecosystem responds differently taking into account 1) time of day, 2) time zone, 3) various holiday and/or vacation patterns, and 4) ever-occurring major newsworthy events. The information sharing model was introduced in the author's Ph.D. dissertation<sup>3</sup>.

The Internet navigational problem led to a decentralized approach for retrieving Web pages located throughout the Internet which combines adaptive mechanisms and policies based on honeybee foraging strategies incorporated into the HRD system to adaptively (and continuously) solve the network routing problem. The network routing problem requires shortest path routing that minimizes "hops" between



the source and randomly chosen TUCs and/or ISPs. Network factors that must be considered are connection requirements (end-to-end delay, delay variation, mean rate) and network conditions such as the self-similarity (fractal behavior) resulting from Web users and file transmissions. The HRD system objectives include maximization of resource utilization and overall LAN, WAN (wide area networks), and Internet (LAN + WAN) throughput. Simultaneously, the HRD system attempts to minimize rejected request packets and guarantee quality of service (QoS). Within each TUC, RS statistics can be used to measure the degree of self-similarity between the HRD system and each website hosted by the TUC.

Self-similarity for a time series dataset<sup>1,5</sup> has been defined using the aggregate sum of  $m - 1$  time measurements over non-overlapping blocks of size  $m$ . The aggregate sum,  $X^{(m)} = (X_k^{(m)}; k = 1, 2, 3, \dots)$ , was rescaled by a factor of  $m^H$  to approximate a zero-mean, stationary time series,  $X = (X_t; t = 1, 2, 3, \dots)$ . Self-similarity exists in the data traffic associated with determining a customized route if  $X^{(m)}$  and  $X$  have the same distribution

$$X_t = m^{-H} \sum_{i=t_m-m+1}^{t_m} X_i \text{ for all } m \in N \quad (1)$$

The presence of a distributional self-similar time series implies that the *autocorrelation function*

$$r(k) = E[(X_t - \mu)(X_{t+k} - \mu)]/\sigma^2 \quad (2)$$

for  $X$  and  $X^m$  differs by a factor of  $m^H$ . The self-similarity parameter  $H = 1 - \frac{\beta}{2}$  is meaningful when  $X = X^{(m)}$ .

The standard usage of these statistical methods<sup>1</sup> are plots for time variance, the RS statistics of the self-similar dataset, the power spectrum frequency, and the Whittle estimator—a measure of the true underlying level of self-similarity. The slow decay variance for a self-similar time series can be plotted using a *variance-time plot*. This log-log plot can be generated by plotting the variance of  $X^{(m)}$  against  $m$ . RS statistic plots can be used to show that self-similar datasets grow according to a power law where the exponent  $H$  is the function of the number of points within each dataset. It should be noted that RS statistic plots are normally used to monitor the self-similar nature of Web traffic for a single site, as opposed to simultaneous monitoring of multiple sites used in this adaptation of the honeybee information search strategies.

## References

1. W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, Proc. ACM SIG-Comm '93, 1-11 (1993).
2. S. Dill, R. Kumar, K. McCurley, S. Rajagopalan, D. Sivakumar, and A. Tomkins, ACM TOIT 2(3), 205-223 (2002).
3. R.L. Walker, Ph.D. Dissertation, Univ. Calif. Los Angeles (2003).
4. T. Berners-Lee, R. Cailliau, J. Groff, and B. Pollermann, Electronic Networking: Research, Applications and Policy 1(1), 74-82 (1992).
5. M.E. Crovella and A. Bestavros, IEEE/ACM Trans. Networking 5(6), 835-846 (1997).

# SIGNAL AND IMAGE PROCESSING WITH FRACLAB

JACQUES LÉVY VÉHEL

*Projet Fractales, INRIA Rocquencourt, 78153 Le Chesnay Cedex, France*

PIERRICK LEGRAND

*IRCCyN, 1 rue de la Née, 44321 Nantes, France*

*e-mails: {vehel,legrand}@irccyn.ec-nantes.fr*

Fractal and multifractal tools have found a large number of applications in recent years. They are increasingly used in areas including astronomy, medical image/signal processing, telecommunications, finance, speech processing, etc... With the spread of fractal analysis in such diverse fields, it seems important that researchers and practitioners willing to make use of fractal tools dispose of a stable set of methods for computing, e.g., fractional dimensions or multifractal spectra. Such methods should be both thoroughly tested and up-to-date, so that they may serve as a benchmark to compare approaches and results. We present in this paper **FracLab**, an *open* and *free* software toolbox that has been developed to serve as such a benchmark. **FracLab** currently contains roughly 800 routines that can be tested and enhanced by the community, and may be used as a reference in various situations. A second aim of **FracLab** is related to a recent evolution in the use of fractal analysis: It has been realized that it is often beneficial to apply fractal tools to arbitrary (i.e. "non-fractal") signals. The best known example is fractal image compression based on IFS theory, as popularized in<sup>1</sup>: IFS-based compression allows to process any kind of images, without an assumption of "fractality". This is also the point of view adopted in **FracLab**: **FracLab** performs *fractal processing* of signals, rather than processing of *fractal signals*. This approach should not be too surprising: Just as, e.g., gradient-based algorithms are often successfully applied for image segmentation even when there are no mathematical or physical reasons for the original signal to possess an ordinary derivative, a fractal analysis may yield new insights for "non-fractal" data. **FracLab** proposes to use fractal analysis in exactly the same way as other mathematical tools are used in everyday signal processing: Under certain assumptions, one may always estimate a gradient from discrete data (for instance *via* a model). In the same way, **FracLab** computes fractional dimensions or multifractal spectra by making adequate assumptions (e.g. that the underlying continuous signal belongs to a parametric class). From a general point of view, fractal analysis with **FracLab** will be of interest when some relevant information is carried in the irregular part of the observations. An example is radar imaging. Such data are difficult to process because of the presence of a specific noise, the *speckle*. However, speckle is not pure noise, but rather a genuine part of the signal, caused by the interferometric nature of radar images. Furthermore, speckle, which is the irregular part of the signal, contains information which is essential about the imaged region. This information is well analyzed with the help of fractal tools (see<sup>3</sup> for more information).

We hope that **FracLab** will help disseminate the use of fractal tools in the processing of irregular but arbitrary signals. This will allow to discover new situations

where fractal analysis yields an interesting alternative to classical signal processing.

We now describe briefly the main functionalities of **FracLab**. **FracLab** may be approached from two points of view: 1- Computation of various fractal parameters and synthesis of fractal signals. 2-Signal and Image processing. In order to make **FracLab** user-friendly, a graphic interface is provided.

**Synthesis of fractal signals.** Two types of signals can be generated : measures or functions. Measures are interesting when one needs to take into account the resolution in an explicit way. For both measures and signals, either deterministic or stochastic data may be generated. **FracLab** allows to synthesize a substantial subset of all classical fractal models described in the literature : 1D/2D fractional Brownian motions, multifractional Brownian motions, (generalized) Weierstrass functions, Lévy motions, (wavelet based)  $1/f$  process<sup>6</sup>, lacunary wavelet series, 1D/2D random multiplicative measures, generalized Riesz products, ...

**Fractal and Multifractal Analysis.** The most basic parameters that can be computed are of course fractional dimensions. In the current implementation of **FracLab**, the  $\text{box}^2$  and *regularization*<sup>5</sup> dimensions are available. In many applications in signal processing, one is more interested in local characterizations of the data. *Hölder exponents* are then more relevant. A specific set of tools allows to estimate both pointwise and local exponents using various methods. In addition, *correlation exponents* may be computed, as well as *2-microlocal exponents*. *Large deviation* and *Legendre Multifractal Spectra* may be computed through various estimations procedures. Finally, **FracLab** allows to test the *Lévy-stability* of a given process and to estimate the associated relevant parameters.

We emphasize the fact that **FracLab** computes fractal exponents through a robust procedure capable of estimating *inferior limits* and not only *plain limits*. This is useful in many real-world applications, where the exponents are not well-defined.

**Signal and Image Processing.** **FracLab** allows to perform *segmentation* of both signals and images. In the former case, a modelling based on a generalization of IFS, called *weakly self affine functions*, is used. Images are segmented into edges/regions of given regularity through multifractal analysis. It is also possible to *regularize* and *denoise* 1D/2D data using various methods based on Hölder regularity analysis or multifractal analysis. Finally, one may *interpolate* 1D/2D data in a such a way that the evolution of various fractal features are controlled in the process.

A few dozens of research groups are known to use **FracLab** at this time.

**FracLab** may be downloaded freely at <http://fractales.inria.fr> or [www.irccyn.ec-nantes.fr/hebergement/FracLab/](http://www.irccyn.ec-nantes.fr/hebergement/FracLab/). More details on **FracLab** may be found in<sup>4</sup>.

## References

1. M. BARNSELY, *Fractals Everywhere*, Academic Press, New-York, 1988.
2. K.J. FALCONER, *Fractal Geometry: Mathematical Foundations and Applications*, John Wiley, New York, 1990.
3. J. LÉVY VÉHEL, Signal enhancement based on Hölder regularity analysis, IMA Volumes in Mathematics and its Applications, Volume 132, pp. 197-20, 2002.
4. J. LÉVY VÉHEL, P. LEGRAND, Signal Processing with FracLab, Inria Res. Rep., 2003.
5. F. ROUEFF, J. LÉVY VÉHEL, A regularization approach to fractional dimension estimation, Fractals conference, Malta, 1998.
6. G.W. WORNELL, Wavelet-Based Representation for the  $1/f$  Family of Fractal Processes, Proc. IEEE, 81 (10), pp 1428-1450, 1993.

## Index

- Allegrini P., 119  
 Andrade Jr. J. S., 47  
 Arvia A. J., 267  
 Ayad G., 315  
  
 Bari A., 315  
 Betancourt-Mar J. A., 311  
 Bryksina N. A., 313  
 Burkle-Elizondo G., 309  
  
 Chiba N., 57  
  
 Da Silva H. F., 47  
 Daughtry C., 317  
 Do M.-T., 189  
 Drakopoulos V., 255  
  
 Elouafi I., 315  
  
 Fisher K., 199  
 Fleury V., 279  
 Fuentes-Larios A. G., 309  
 Fujimoto T., 57  
  
 Gonzalez-Andujar J. L., 315  
 Gorenflo R., 35  
 Gouyet J.-F., 279  
 Grigolini P., 119  
  
 Haase M., 69  
 Henrique E. A., 47  
 Honjo H., 233  
  
 Jones O. D., 165  
  
 Kakos A., 255  
 Karimova L. M., 91  
 Katsuragi H., 233  
 Kaul M., 317  
 Kimmerer W., 199  
 Kuandykov Y. B., 91  
  
 Lévy Véhel J., 189, 321  
 Last W. M., 313  
 Legrand P., 189, 321  
  
 Lehle B., 69  
 Li J., 213  
 Lydon J., 317  
  
 Mainardi F., 35  
 Makarenko N. G., 91  
 Mandelbrot B. B., 1  
 Marchiano S. L., 267  
 Martin A., 315  
 Martyn T., 79  
 Mora A., 69  
  
 Nachit M., 315  
 Navascues M. A., 143  
 Nekka F., 213  
 Nguyen M. B., 279  
 Nikolaou N., 255  
 Novak M. M., 1, 91  
  
 Pachepsky L., 317  
 Palatella L., 119  
 Pasquale M. A., 267  
 Prusinkiewicz P., 103  
  
 Sakaguchi H., 301  
 Sala N., 177  
 Sapoval B., 47  
 Saucier A., 291  
 Scalas E., 35  
 Sebastian M. V., 143  
 Stauffer D., 131  
 Suárez-Domínguez E. J., 311  
 Sugino D., 233  
 Sun H., 243  
  
 Tikhomolov E., 155  
 Tokunaga S., 301  
  
 Valdez-Cepeda R. D., 309  
  
 Walker R. L., 319  
 Walthall Ch., 317  
 Woloszyn P. W., 221  
  
 Xie H., 243